

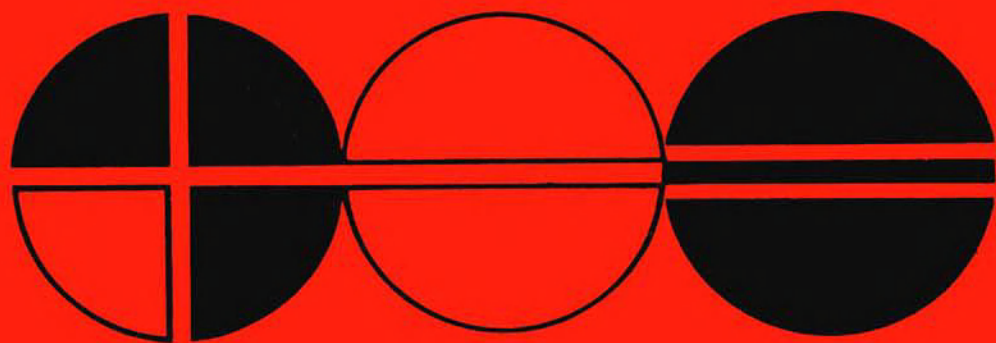
NORTH-HOLLAND

MATHEMATICS STUDIES

61

Nonlinear Problems: Present and Future

ALAN BISHOP
DAVID CAMPBELL
BASIL NICOLAENKO
Editors



NORTH-HOLLAND

NONLINEAR PROBLEMS: PRESENT AND FUTURE

This Page Intentionally Left Blank

Nonlinear Problems: Present and Future

Proceedings of the First Los Alamos Conference
on Nonlinear Problems,
Los Alamos, NM, U.S.A., March 2-6, 1981

Edited by

**ALAN BISHOP
DAVID CAMPBELL
BASIL NICOLAENKO**

*Los Alamos National Laboratory
Los Alamos, New Mexico, U.S.A.*



1982

© North-Holland Publishing Company 1982

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 0 444 86395 8

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

Sole distributors for the U.S.A. and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE
NEW YORK, N.Y. 10017

Library of Congress Cataloging in Publication Data

Los Alamos Conference on Nonlinear Problems
(1st : 1981)
Nonlinear problems.

(North-Holland mathematics studies ; 61)

1. Nonlinear theories--Congresses. 2. Mathematical physics--Congresses. I. Bishop, A.R. (Alan R.) II. Campbell, David, 1944 July 23-
III. Nicolaenko, Basil, 1943- . IV. Title.
V. Series.
QC20.7.N6L67 1981 530.1'5 82-3504
ISBN 0-444-86395-8 AACR2

PRINTED IN THE NETHERLANDS

PREFACE

Recognizing the growing awareness of common problems and answers in the nonlinear sciences, the Los Alamos Center for Nonlinear Studies (CNLS) was established by the Laboratory Director in October 1980, to coordinate interdisciplinary studies, to strengthen ties among Laboratory researchers and with the academic community, and to interface between basic and applied areas. The breadth of research problems at Los Alamos and the variety of technical expertise reflected in its staff offers a fertile environment for the CNLS. But from its inception, the Center was intended not for Los Alamos alone, but also as an international resource for the entire nonlinear community.

To realize this intention, the CNLS, through its Chairman, Alwyn C. Scott, has developed several continuing modes of operation:

- identifying broad themes for intensive interdisciplinary research (currently, these themes are nonlinear phenomena in reactive flows and chaos and coherence in physical systems);
- coordinating an active visitor program, including both guest lecture series and longterm collaborative visits;
- sponsoring technical workshops aimed at bringing experts together to expound recent results and delineate future directions in specific nonlinear problems (recent workshops have included "Adaptive Grid Methods," "Coupled Nonlinear Oscillators," and "Solitons and the Bethe Ansatz"); and
- hosting an annual international conference on some topical area of nonlinear science.

In the last category, the major event for the CNLS in its first year was an international conference on 'Nonlinear Problems: Present and Future' held at the Los Alamos National Security and Resources Study Center, Los Alamos, March 2-6, 1981, chaired by Mark Kac and Stanislaw Ulam. This volume contains the edited proceedings of that conference. We are honored to dedicate the proceedings to Fermi, Pasta, and Ulam, distinguished Los Alamos alumni who have set a tradition of excellence to which the CNLS must aspire. [We have all been saddened by the death of John Pasta since the Conference took place and hope that this volume will be accepted as our small tribute to his imaginative career.]

As befits an inaugural conference, a very wide spectrum of topics were represented ranging from pure mathematics, through numerical methods, to sophisticated experiments on fluids and solids. More specialized meetings are anticipated in future years including some of the many topics that it was impossible to cover in 1981. However, the deliberately interdisciplinary atmosphere of the inaugural meeting truly reflected an exciting stage of development of nonlinear science as a unified subject. The provocative title attracted well over two hundred participants and a distinguished list of speakers, many of whom succeeded admirably in overcoming scientific language barriers and generating a broad interest in their fundamental problems. Four major topics were represented through survey lectures and workshop activities: turbulence in plasmas and fluids (both the onset and fully-developed turbulence); nonlinearity in field theory and in low-dimensional solids; reaction-diffusion processes; and new methods in nonlinear mathematics. As described in detail in the Contents, we have preserved these divisions in the Proceedings, supplementing invited papers with a small number of relevant contributed ones. It is our firm impression that these articles, through survey and original work, represent the cutting edges of several important areas of nonlinearity. We hope they will be valuable reading for novices and experts alike.

We think that all those who attended the Conference will remember it for its stimulation and unobtrusive organization. This crucial combination could not have been achieved without the advice and support of our colleagues in the CNLS. The Director and his efficient staff provided every conceivable help in coordinating the splendid Laboratory facilities. Mark Kac and Stanislaw Ulam were supportive conference chairman and Stanislaw Ulam graciously agreed to give a nostalgic after-dinner speech on the FPU problem which was admired by all. No conference is better than its secretary. In Janet Gerwin we had a secretary whose competence was apparent at every stage before, during, and after the conference. All three organizers are immeasurably indebted to her for her skill in the face of the continual crises! We are also happy to thank Janet, Chris Davis, Frankie Gomez, Mary Plehn, and Kate Procknow for their skillful assistance in preparing these proceedings. Last but not least we must thank our publishers for their excellent cooperation and every conference attendee for joining us in this celebration of nonlinear physics.

Los Alamos

A. R. Bishop
D. K. Campbell
B. Nicolaenko

DEDICATION

It started with an experiment--the new kind in which the same instrument, the new computer, both creates and probes an idealization of the real world.

The purpose, quite modest, was to test what seemed beyond doubt, namely, that in a nonlinear discretized string the energy initially concentrated in one vibrational mode ultimately distributes itself among all modes. The result, first announced in a Los Alamos report, was however, startlingly different: after an initial tendency toward equipartition, the energy flowed back to the initial mode.

Thus a new chapter of nonlinear science began and much of its spectacular growth in the past quarter of a century is directly or indirectly traceable to the pioneering experiment with the nonlinear string.

It is therefore fitting that this Conference marking the creation of the Los Alamos Center for Nonlinear Studies be dedicated to the authors of that historic 1955 report: Fermi, Pasta, and Ulam.

Los Alamos

M. Kac
N. Metropolis

This Page Intentionally Left Blank

TABLE OF CONTENTS

PREFACE	v
DEDICATION	vii
PART I : NEW METHODS AND RESULTS IN NONLINEAR ANALYSIS	
Optimal Control of Non-Well-Posed Distributed Systems and Related Nonlinear Partial Differential Equations J.-L. LIONS	3
Compactness and Topological Methods for Some Nonlinear Variational Problems of Mathematical Physics P.L. LIONS	17
On Yang-Mills Fields I.M. SINGER	35
Gauge Theories for Soliton Problems D.H. SATTINGER	51
The Inverse Monodromy Transform is a Canonical Transformation H. FLASCHKA and A.C. NEWELL	65
Numerical Methods for Nonlinear Differential Equations J.M. HYMAN	91
Limit Analysis of Physical Models M. FREMOND, A. FRIAA, and M. RAUPP	109
Viscosity Solutions of Hamilton-Jacobi Equations M.G. CRANDALL	117
Bifurcation of Stationary Vortex Configurations J.I. PALMORE	127
Exact Invariants for Time-Dependent Nonlinear Hamiltonian Systems H.R. LEWIS and P.G.L. LEACH	133
Isolating Integrals in Galactic Dynamics and the Character of Stellar Orbits W.I. NEWMAN	147
PART II : NONLINEARITY IN FIELD THEORIES AND LOW DIMENSIONAL SOLIDS	
Physics in Few Dimensions V.J. EMERY	159
Solution of the Kondo Problem N. ANDREI	169
Kinks of Fractional Charge in Quasi-One Dimensional Systems J.R. SCHRIEFFER	179

Theoretically Predicted Drude Absorption by a Conducting Charged Soliton in Doped-Polyacetylene M.J. RICE	189
Polarons in Polyacetylene A.R. BISHOP and D.K. CAMPBELL	195
Light Scattering and Absorption in Polyacetylene S. ETEMAD and A.J. HEEGER	209
Quasi-Solitons: A Case Study of the Double Sine-Gordon Equation P. KUMAR and R.R. HOLLAND	229
Classical Field Theory with $Z(3)$ Symmetry H.M. RUCK	237
PART III : REACTION-DIFFUSION PROCESSES	
Some Characteristic Nonlinearities of Chemical Reaction Engineering R. ARIS	247
Propagating Fronts in Reactive Media P.C. FIFE	267
PART IV : NONLINEAR PHENOMENA IN FLUIDS AND PLASMAS	
Regularity Results for the Equations of Incompressible Fluids Mechanic at the Brink of Turbulence C. BARDOS	289
Finite Parameter Approximative Structure of Actual Flows C. FOIAS and R. TEMAM	317
The Role of Characteristic Boundaries in the Asymptotic Solution of the Navier-Stokes Equations F.A. HOWES	329
Some Approaches to the Turbulence Problem J. MATHIEU, D. JEANDEL, and C.M. BRAUNER	335
Instability of Pipe Flow A.T. PATERA and S.A. ORSZAG	367
Some Formalism and Predictions of the Period-Doubling Onset of Chaos M.J. FEIGENBAUM	379
Tricritical Points and Bifurcations in a Quartic Map S.-J. CHANG, M. WORTIS, and J.A. WRIGHT	395
Recent Experiments on Convective Turbulence J.P. GOLLUB	403
Experimental Observations of Complex Dynamics in a Chemical Reaction J.C. ROUX, J.S. TURNER, W.D. McCORMICK, and H.L. SWINNEY	409
Nonlinear Plasma Dynamics Below the Cyclotron Frequency J.M. GREENE	423

TABLE OF CONTENTS

xi

Turbulence and Self-Consistent Fields in Plasmas D. PESME and D.F. DUBOIS	435
Self-Focusing Tendencies of the Nonlinear Schrodinger and Zakharov Equations H.A. ROSE and D.F. DUBOIS	465
Chaotic Oscillations in a Simplified Model for Langmuir Waves P.K.C. WANG	479
INDEX OF CONTRIBUTORS	483

This Page Intentionally Left Blank

PART I
New Methods and Results
in Nonlinear Analysis

This Page Intentionally Left Blank

OPTIMAL CONTROL OF NON WELL POSED DISTRIBUTED SYSTEMS AND RELATED NON LINEAR PARTIAL DIFFERENTIAL EQUATIONS

Jacques-Louis LIONS

Collège de France and INRIA

Various practical problems lead to the question of optimal control of non well posed systems - such as non linear unstable systems. We begin with simple linear - quadratic examples of non well posed evolution systems ; the optimality system leads to new non linear Partial Differential Equations of Ricatti's type. We study then non linear parabolic unstable systems, with a distributed or a boundary control, with or without constraints. The optimality system is given.

INTRODUCTION

1. Let \mathcal{A} be a partial differential operator, linear or not, of evolution or of stationary type. In the usual theory of optimal control of distributed systems, the state equation is given, in a formal manner, by

$$\mathcal{A}y = \mathcal{B}v \quad (1)$$

where v denotes the control variable (or function) and \mathcal{B} is an operator which can be thought of as giving boundary conditions ; in (1) one has to add initial conditions if \mathcal{A} is of evolution type.

In the usual theory one assumes that, given v in a suitable Banach space U , equation (1) subject to appropriate boundary and initial conditions, admits a unique solution denoted by $y(v)$; $y(v)$ is the state of the system ; $y(v)$ belongs to a space Y when v spans U .

Then the cost function is given by

$$J(v) = \phi(y(v)) + \psi(\|v\|_U) \quad (2)$$

where ϕ is a continuous functional from $Y \rightarrow \mathbb{R}$, where $\|v\|_U$ denotes the norm of v in U and where $\lambda \rightarrow \psi(\lambda)$ is continuous for $\lambda \geq 0$, $\psi(0) = 0$, $\psi(\lambda) \rightarrow +\infty$ as $\lambda \rightarrow +\infty$.

If U_{ad} denotes a (suitable) subset of U , then the problem is to find

$$\inf J(v), \quad v \in U_{ad}. \quad (3)$$

If (3) admits a solution u , one of the main questions is then to find a set of necessary (or necessary and sufficient) conditions for characterizing u , i.e. to find the optimality system. For these questions we refer to J.L. LIONS [1][2][3] and to the Bibliography therein.

2. A slightly different situation can occur if the functional ϕ in (2) is not defined on the whole space Y . Then one has to introduce new functional spaces.

Let us give an example. Let Ω be a bounded open set of \mathbb{R}^3 with boundary Γ ; we consider the state equation

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y = v(t)\delta(x-b) & \text{in } Q = \Omega \times]0, T[\\ y = 0 & \text{on } \Sigma = \Gamma \times]0, T[, \\ y(x, 0) = 0 \end{cases} \quad (4)$$

where $\delta(x-b)$ denotes the Dirac measure at point $b \in \Omega$.

Given $v \in L^2(0, T)$, problem (4) admits a unique weak solution (cf. J.L. LIONS and E. MAGENES [1]) $y(v) \in L^2(Q)$.

Let us consider now the cost function

$$J(v) = \int_{\Omega} [y(x, T; v) - z_d]^2 dx + N \int_0^T v^2 dt, \quad (5)$$

where z_d is given in $L^2(\Omega)$ and $N > 0$. In general, for $v \in L^2(0, T)$, $y(\cdot, T; v)$ is defined but as an element of $H^{-1}(\Omega)$ (Sobolev space of order -1) and not an element of $L^2(\Omega)$. Therefore (5) does not make sense for $v \in L^2(0, T)$. One has then to restrict J to those v 's such that

$$v \in L^2(0, T) \text{ and } y(\cdot, T; v) \in L^2(\Omega). \quad (6)$$

This defines a Hilbert space (when provided with the norm $(\int_0^T v^2 dt + \int_{\Omega} y(x, T; v)^2 dx)^{1/2}$ say U , and if

$$U_{ad} \subset U, \quad (7)$$

we consider again problem (3). In order to proceed it is necessary to study U , not only to make things more precise but also because the dual U' of U is needed for writing the optimality system. One verifies that U coincides with the set of those v 's in $L^2(0, T)$ such that

$$\int_0^T \int_0^T (2T - (t+s))^{-3/2} v(t)v(s) dt ds < \infty. \quad (8)$$

For questions of this type we refer to J.L. LIONS [4][5], J. SIMON [1].

3. A third situation can occur when (1) is not a well posed problem. Equations (1) which have a physical interest and which lead to non well set problems arise in unstable phenomena, in situations where we have bifurcations - cf. J.P. KERNEVEZ J.L. LIONS and D. THOMAS [1]. One has then to change significantly the point of view. One considers the set of v and z such that

$$v \in U, \quad z \in Y \quad (9)$$

$$Az = Bv. \quad (10)$$

Then one considers the cost function

$$J(v, z) = \phi(z) + \psi(\|v\|_U) \quad (11)$$

and one looks for

$$\inf J(v, z), \quad v, z \text{ subject to } (9) (10) \quad (12)$$

with the possible added constraint

$$v \in U_{ad} \quad \blacksquare \quad (13)$$

As an example (without physical interest) we consider

$$\begin{cases} \frac{\partial z}{\partial t} + \Delta z = v & \text{in } Q \\ v, z \in L^2(Q) \end{cases} \quad (14)$$

with the conditions

$$z(x, 0) = 0, \quad z = 0 \text{ on } \Sigma \quad (15)$$

(one can prove that conditions (15) do make sense ; cf. Section 1 below). Let the cost function be given by

$$J(v, z) = \|z - z_d\|_{L^2(Q)}^2 + N \|v\|_{L^2(Q)}^2 \quad (16)$$

and let U_{ad} be a closed convex subset of $L^2(Q)$ such that the set of those v, z 's such that $v \in U_{ad}$ and (14) (15) hold true is not empty. Then

$$\inf J(v, z), \quad v \in U_{ad}, \quad v, z \text{ satisfy (14) (15)} \quad (17)$$

admits a unique solution $\{u, y\}$. \blacksquare

Returning to the general case, we want to find an optimality system for these problems of optimal control.

4. We consider in this paper three (of the many) situations of such problems.

In Section 1 we consider a system of type (14) but which is also non well posed for $t < T$, namely

$$\frac{\partial z}{\partial t} + m(t) \Delta z = v, \quad (18)$$

with $m > 0$ (resp. < 0) near 0 (resp. near T).

Decoupling the optimality system leads to apparently new nonlinear Partial Differential equations.

In Section 2 we consider unstable systems governed by

$$\frac{\partial z}{\partial t} - \Delta z - z^3 = v \quad (19)$$

(or with boundary control).

Other situations are indicated in J.L. LIONS [5], such as the case of elliptic systems which can be controlled by Cauchy data on part of the boundary.

Problems of optimum design where again the state equation is not well set will be studied elsewhere.

5. Existence problems for not necessarily well set problems (such as Navier-Stokes equations in space dimension equal 3) have been studied by A.V. FOURSICOV [1] ; this author does not consider the optimality system.

The optimality system for problem (17) involves new functional spaces (of distributions of infinite order) ; we refer to P. RIVERA [1].

1. New non linear Partial Differential equation of Riccati's type.

1.1. Setting of the problem.

We consider couples $\{v, z\}$ such that

$$v, z \in L^2(Q) \times L^2(Q), \quad Q = \Omega \times]0, T[, \quad (1.1)$$

and

$$\frac{\partial z}{\partial t} + m(t) z = v \text{ in } Q, \quad (1.2)$$

$$z(x, 0) = 0 \text{ in } \Omega, \quad (1.3)$$

$$z = 0 \text{ on } \Sigma = \Gamma \times]0, T[. \quad (1.4)$$

In (1.2) m denotes a continuous function with a graph as represented on Fig. 1.

Conditions (1.3) (1.4) make sense.

Let us check it for (1.3); it follows from (1.1) (1.2) that

$$\begin{cases} z \in L^2(0, T; L^2(\Omega)), \\ \frac{\partial z}{\partial t} \in L^2(0, T; H^{-2}(\Omega)) \end{cases} \quad (1.5)$$

(where $H^{-2}(\Omega)$ = dual of $H_0^2(\Omega)$, $H_0^2(\Omega) = \{\phi | \phi, \frac{\partial \phi}{\partial x_i}, \frac{\partial^2 \phi}{\partial x_i \partial x_j} \in L^2(\Omega), \phi = 0, \frac{\partial \phi}{\partial x_i} = 0 \text{ on } \Gamma\}$.) ;

it follows from (1.5) and from standard results that z is continuous form $[0, T] \rightarrow H^{-1}(\Omega)$ so that (1.3) makes sense. One checks by similar techniques that (1.4) makes sense.

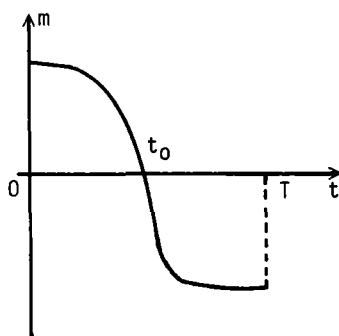


Fig. 1

Of course (1.2) (1.3) (1.4) is a non well posed problem. Moreover the system is also non well posed in the backward time direction; if we replace (1.3) by $z(x, T) = 0$ in Ω , the corresponding problem is also non well posed.

The cost function that we consider is given by

$$J(v, z) = \|z - z_d\|_{L^2(Q)}^2 + N \|v\|_{L^2(Q)}^2, \quad (1.6)$$

z_d given in $L^2(Q)$, $N > 0$.

We consider the "no-constraints" problem, namely

$$\inf J(v, z), \quad v, z \text{ subject to (1.1)...(1.4)}. \quad (1.7)$$

1.2. Optimality system.

It is a simple matter to check that problem (1.7) admits a unique solution u, y . We want to characterize u, y ; this characterization is given by the following result : there exists a unique set $\{u, y, p\}$ of functions such that

$$\frac{\partial y}{\partial t} + m(t)\Delta y = u, \quad -\frac{\partial p}{\partial t} + m(t)\Delta p = y - z_d, \quad (1.8)$$

$$y = p = 0 \text{ on } \Sigma, \quad (1.9)$$

$$y(x, 0) = 0, \quad p(x, T) = 0 \quad (1.10)$$

$$y, p \in L^2(Q), \quad (1.11)$$

$$p + Nu = 0 \quad (1.12)$$

One can of course eliminate u by (1.8) (1.12). The system in u, y, p is called the optimality system ; p is a Lagrange multiplier.

The proof of this result can be obtained as follows ; one considers the penalized problem

$$J_\varepsilon(v, z) = \|z - z_d\|_{L^2(Q)}^2 + N \|v\|_{L^2(Q)}^2 + \frac{1}{\varepsilon} \left\| \frac{\partial z}{\partial t} + m(t)\Delta z - v \right\|_{L^2(Q)}^2 \quad (1.13)$$

where $v \in L^2(Q)$, $z \in L^2(Q)$, $\frac{\partial z}{\partial t} + m(t)\Delta z \in L^2(Q)$ and $z(x, 0) = 0$; $z = 0$ in Σ , and where $\varepsilon > 0$ is "small".

Then

$$\inf J_\varepsilon(v, z) = J_\varepsilon(u_\varepsilon, y_\varepsilon). \quad (1.14)$$

One defines p_ε by

$$p_\varepsilon = -\frac{1}{\varepsilon} \left(\frac{\partial y_\varepsilon}{\partial t} + m\Delta y_\varepsilon - u_\varepsilon \right). \quad (1.15)$$

One verifies that

$$\begin{cases} -\frac{\partial p_\varepsilon}{\partial t} + m(t)\Delta p_\varepsilon = y_\varepsilon - z_d \text{ in } Q, \\ p_\varepsilon(x, T) = 0, \quad p_\varepsilon = 0 \text{ on } \Sigma \end{cases} \quad (1.16)$$

and that

$$p_\varepsilon + Nu_\varepsilon = 0 \text{ in } Q. \quad (1.17)$$

It follows from (1.13) and (1.17) that

$$u_\varepsilon, y_\varepsilon, p_\varepsilon \text{ remain, as } \varepsilon \rightarrow 0, \text{ in a bounded set of } (L^2(Q))^3, \quad (1.18)$$

and that

$$\frac{\partial y_\varepsilon}{\partial t} + m(t)\Delta y_\varepsilon - u_\varepsilon = \sqrt{\varepsilon} g_\varepsilon, \quad \|g_\varepsilon\|_{L^2(Q)} \leq C \quad (1.19)$$

Therefore one can pass to the limit in $\varepsilon \rightarrow 0$. One verifies that $u_\varepsilon, y_\varepsilon \rightarrow u, y$ in $L^2(Q) \times L^2(Q)$ and that u, y, p satisfy the optimality system.

One can write a variational formulation for the problem, as follows : p is the solution of

$$\left(-\frac{\partial p}{\partial t} + m(t)\Delta p, -\frac{\partial q}{\partial t} + m(t)\Delta q \right) + \frac{1}{N} (p, q) = -(z_d, -\frac{\partial q}{\partial t} + m\Delta q) \quad \forall q \quad (1.20)$$

(where $(p, q) = \int_Q p q \, dx \, dt$) where $q \in L^2(Q)$, $-\frac{\partial q}{\partial t} + m(t)\Delta q \in L^2(Q)$, $q(x, T) = 0$,

$q = 0$ on Σ . ■

1.3. We are now going to show how to "uncouple" the optimality system.

The technique is similar to the one in J.L. LIONS [1] but one deals now with weak solutions, due to the fact that the "state equation" is not well posed.

One considers a problem analogous to (1.8)...(1.12) but in the interval $]s, T[$, $0 < s < T$:

$$\begin{cases} \frac{\partial \phi}{\partial t} + m\Delta \phi + \frac{1}{N} \psi = 0, & -\frac{\partial \psi}{\partial t} + m\Delta \psi = \phi - z_d, & \text{in } \Omega \times]s, T[, \\ \phi(x, s) = h(x) \text{ in } \Omega, & \psi(x, T) = 0 \text{ in } \Omega, \\ \phi = \psi = 0 & \text{on } \Gamma \times]s, T[. \end{cases} \quad (1.21)$$

In (1.21) h is given in, say, the space of C^∞ smooth functions with compact support. Problem (1.21) is the optimality system for a problem entirely analogous to (1.6) (1.7) but with $\Omega \times]0, T[$ replaced by $\Omega \times]s, T[$ and with $z(x, 0) = 0$ replaced by $\phi(x, s) = h(x)$. Therefore (1.21) admits a unique solution and

$$\psi(x, s) \text{ is uniquely defined.} \quad (1.22)$$

The space where $\psi(\cdot, s)$ belongs depends whether $s < t_0$ or $s > t_0$. If $s < t_0$, one deals with a well set system, and $\psi(\cdot, s) \in H_0^1(\Omega)$; if $s > t_0$, one has to work with weak solutions and $\psi(\cdot, s) \in H^{-1}(\Omega)$.

We have :

$$\psi(\cdot, s) = P(s)h + r(s) \quad (1.23)$$

where $P(s)$ is a linear operator; one has

$$\begin{cases} P(s) \in \mathcal{L}(H^{-1}(\Omega); H^{-1}(\Omega)) & \text{if } s \leq t_0, \\ P(s) \in \mathcal{L}(H^{-1}(\Omega); H_0^1(\Omega)) & \text{if } s > t_0. \end{cases} \quad (1.24)$$

If we take in (1.21) $h = y(\cdot, s) = y(s)$ then ϕ, ψ = restriction of y, p to $\Omega \times]s, T[$, so that (1.23) becomes (changing s into t) :

$$p(t) = P(t)y(t) + r(t). \quad (1.25)$$

Using the L. Schwartz kernel theorem [1], one sees that

$$P(t)h = \int_{\Omega} P(x, \xi, t) h(\xi) d\xi \quad (1.26)$$

where the kernel $P(x, \xi, t)$ is a distribution on $\Omega_x \times \Omega_\xi$.

Using (1.25) into (1.8)...(1.12) one obtains finally that $P(x, \xi, t)$ satisfies

$$-\frac{\partial P}{\partial t} + m(t)(\Delta_x + \Delta_\xi)P + \frac{1}{N} \int_{\Omega} P(x, \zeta, t) P(\zeta, \xi, t) d\zeta = \delta(x - \xi), \quad (1.27)$$

$$P(x, \xi, T) = 0 \quad (1.28)$$

and

$$P(x, \xi, t) = P(\xi, x, t) \quad \forall x, \xi \in \Omega \times \Omega. \quad (1.29)$$

The Boundary conditions are of Dirichlet type, in the usual sense for $t_0 < t < T$

and in a very weak sense for $0 < t < t_0$ (corresponding to the fact that $y = 0$ on $\Gamma \times]0, t_0[$ in a very weak sense). ■

1.4. Formal Remarks.

Remark 1.1 : Equation (1.27) is a non linear equation of Riccati's type. The operator is well set for $t_0 < t < T$; when $t < t_0$, the linear part of the operator becomes non well set (in the backward time direction) but the whole operator remains well set in a larger space; $\mathcal{L}(H^{-1}(\Omega) ; H_0^1(\Omega))$ is replaced by $\mathcal{L}(H^{-1}(\Omega) ; H^{-1}(\Omega))$. ■

Remark 1.2 : Of course one can consider functions $m(t)$ which change sign in $[0, T]$ an arbitrary number of times. Assume for instance that $m(t)$ is given on \mathbb{R} , with period 1, and changing sign in $[0, 1]$ at least once, and such that $m > 0$ near 0. We define

$$m_\varepsilon(t) = m(t/\varepsilon). \quad (1.30)$$

All what has been said is valid with m replaced by m_ε as defined by (1.30). Therefore there exists $P_\varepsilon(x, \xi, t)$ solution of (1.27) (1.28) (1.29) where m is replaced by m_ε .

Finally when $\varepsilon \rightarrow 0$, m_ε should be replaced by

$$\mu = \int_0^1 m(t) dt \quad (1.31)$$

and one obtains (formally) for the limit of P_ε :

$$-\frac{\partial P}{\partial t} + \mu(\Delta_x + \Delta_z)P + \frac{1}{N} \int_\Omega P(x, \xi, t) P(\xi, \xi, t) d\xi = \delta(x - \xi); \quad (1.32)$$

it shows that the properties of regularity of the limit of P_ε depends on the sign of μ ; one has $P \in \mathcal{L}(H^{-1}(\Omega) ; H_0^1(\Omega))$ (resp. $P \in \mathcal{L}(H^{-1}(\Omega) ; H^{-1}(\Omega))$) if $\mu < 0$ (resp. $\mu > 0$).

If $\mu = 0$, $P(x, \xi, t) = \sqrt{N} \left(\text{th} \frac{T-t}{\sqrt{N}} \right) \delta(x - \xi)$. ■

Remark 1.3 : One can also consider the equation

$$\frac{\partial z}{\partial t} + Az = v \quad (1.33)$$

where $Az = -\frac{\partial}{\partial x} \left(x \frac{\partial z}{\partial x} \right)$ on $\mathbb{R} = \Omega$, with $z(x, 0) = 0$ (this is again a non well posed problem). The corresponding Riccati's type equation will be :

$$\begin{cases} -\frac{\partial P}{\partial t} - \frac{\partial}{\partial x} \left(x \frac{\partial P}{\partial x} \right) - \frac{\partial}{\partial \xi} \left(\xi \frac{\partial P}{\partial \xi} \right) + \frac{1}{N} \int_{-\infty}^{+\infty} P(x, \xi, t) P(\xi, \xi, t) d\xi = \delta(x - \xi), \\ P(x, \xi, T) = 0, \\ P(x, \xi, t) = P(\xi, x, t). \end{cases} \quad (1.34)$$

2. - Control of unstable systems.

2.1. Distributed control without constraints.

We are given v, z such that

$$v, z \in L^2(Q) \times L^6(Q), \quad (2.1)$$

$$\begin{cases} \frac{\partial z}{\partial t} - \Delta z - z^3 = v \text{ in } Q, \\ z(x,0) = 0 \text{ in } \Omega, z = 0 \text{ on } \Sigma. \end{cases} \quad (2.2)$$

Remark 2.1 : Given v in $L^2(Q)$, problem (2.2) does not necessarily admit a global solution in time. The solution, defined locally, can blow up. ■

Remark 2.2 : If $z \in L^6(Q)$ then it follows from (2.2) that

$$\frac{\partial z}{\partial t} - \Delta z = v + z^3 \in L^2(Q) \quad (2.3)$$

which implies, with the boundary conditions in (2.2), that

$$z \in H^{2,1}(Q) \text{ i.e. } z, \frac{\partial z}{\partial x_i}, \frac{\partial^2 z}{\partial x_i \partial x_j}, \frac{\partial z}{\partial t} \in L^2(Q). \quad \blacksquare \quad (2.4)$$

The cost function $J(v,z)$ is given by

$$J(v,z) = \frac{1}{6} \|z - z_d\|_{L^6(Q)}^6 + \frac{N}{2} \|v\|_{L^2(Q)}^2 \quad (2.5)$$

where z_d is given in $L^6(Q)$ and where $N > 0$.

In the case without constraint we want to find

$$\inf J(v,z), \quad v, z \text{ subject to (2.1) (2.2)}. \quad \blacksquare \quad (2.6)$$

One has first

$$\text{problem (2.6) admits a solution } \{u, y\}. \quad (2.7)$$

(There is no reason why this solution should be unique). For the proof, we consider a minimizing sequence v_n, z_n . By virtue of (2.5), v_n, z_n remain in a bounded set of $L^2(Q) \times L^6(Q)$. Then $v_n + z_n^3$ remains in a bounded set of $L^2(Q)$ so that z_n remains in a bounded set of $H^{2,1}(Q)$. Therefore one can extract a subsequence, still denoted by v_n, z_n , such that $v_n \rightarrow u$ in $L^2(Q)$ weakly, $z_n \rightarrow y$ in $H^{2,1}(Q) \cap L^6(Q)$ weakly. Since, in particular, $H^{2,1}(Q) \rightarrow L^2(Q)$ is compact, $z_n^3 \rightarrow y^3$ in $L^2(Q)$ weakly and therefore

$$\frac{\partial y}{\partial t} - \Delta y - y^3 = u, \quad y(x,0) = 0, \quad y=0 \text{ on } \Sigma. \quad \blacksquare$$

One can then show that if u, y is a solution of (2.6) then there exists p such that $p \in L^2(Q)$ and

$$p \in W^{2,1;6/5}(Q) \text{ i.e. } p, \frac{\partial p}{\partial x_i}, \frac{\partial^2 p}{\partial x_i \partial x_j}, \frac{\partial p}{\partial t} \in L^{6/5}(Q), \quad (2.8)$$

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y - y^3 = u, & -\frac{\partial p}{\partial t} - \Delta p - 3y^2 p = (y - z_d)^5, \\ y(x,0) = 0, & p(x,T) = 0, \\ y = p = 0 \text{ on } \Sigma, \end{cases} \quad (2.9)$$

$$p + Nu = 0 \text{ in } Q \quad (2.10)$$

For the proof, one considers, as in Section 1.2, the penalized problem

$$J_{\varepsilon}(v, z) = \frac{1}{6} \|z - z_d\|_{L^6(Q)}^6 + \frac{N}{2} \|v\|_{L^2(Q)}^2 + \frac{1}{2\varepsilon} \left\| \frac{\partial z}{\partial t} - \Delta z - z^3 - v \right\|_{L^2(Q)}^2 \quad (2.11)$$

for $v \in L^2(Q)$, $z \in L^6(Q)$, $z \in H^{2,1}(Q)$, $z(0) = 0$, $z|_{\Sigma} = 0$.

We consider $\inf J_{\varepsilon}(v, z)$ and we denote by $v_{\varepsilon}, y_{\varepsilon}$ a solution (which exists !) of this penalized problem. We set

$$-\frac{1}{\varepsilon} \left(\frac{\partial y_{\varepsilon}}{\partial t} - \Delta y_{\varepsilon} - y_{\varepsilon}^3 - u_{\varepsilon} \right) = p_{\varepsilon} \quad (2.12)$$

We have

$$\begin{cases} -\frac{\partial p_{\varepsilon}}{\partial t} - \Delta p_{\varepsilon} - 3y_{\varepsilon}^2 p_{\varepsilon} = (y_{\varepsilon} - z_d)^5 & \text{in } Q, \\ p_{\varepsilon}(x, T) = 0, \quad p_{\varepsilon} = 0 & \text{on } \Sigma, \end{cases} \quad (2.13)$$

and

$$p_{\varepsilon} + Nu_{\varepsilon} = 0. \quad (2.14)$$

Therefore $p_{\varepsilon} = -Nu_{\varepsilon}$ is bounded (as $\varepsilon \rightarrow 0$) in $L^2(Q)$; since y_{ε} is bounded in $L^6(Q)$, it follows that

$$3y_{\varepsilon}^2 p_{\varepsilon} + (y_{\varepsilon} - z_d)^5 \text{ is bounded in } L^{6/5}(Q)$$

so that p_{ε} is bounded in $W^{2,1;6/5}(Q)$, hence the result will follow. ■

2.2. Distributed control with constraints.

Let us take again the situation (2.1)(2.2) of Section 2.1; we add now the constraints

$$v \in U_{ad}, \quad U_{ad} = \text{closed convex subset of } L^2(Q), \quad (2.15)$$

and we make the (strong) assumption :

$$U_{ad} \text{ has a non empty interior.} \quad (2.16)$$

We also assume that the set of $\{v, z\}$ such that (2.1)(2.2)(2.15) take place is not empty.

By a suitable adaptation of an idea of P. RIVERA [1] one can again show the existence of a Lagrange multiplier p for a solution u, y of

$$J(u, y) = \inf J(v, z), \quad v, z \text{ satisfy (2.1)(2.2)(2.15).} \quad (2.17)$$

More precisely, there exists $p \in L^2(Q)$ satisfying (2.8) (2.9) and

$$(p + Nu, v - u) \geq 0 \quad \forall v \in U_{ad}, \quad u \in U_{ad} \quad \blacksquare \quad (2.18)$$

2.3. Boundary control.

Let us assume now that v and z satisfy

$$v \in L^2(\Sigma), \quad z \in L^6(Q) \quad (2.19)$$

$$\begin{cases} \frac{\partial z}{\partial t} - \Delta z - z^3 = 0 \text{ in } Q, \\ \frac{\partial z}{\partial \nu} = v \text{ on } \Sigma, \\ z(x, 0) = 0 \text{ in } \Omega. \end{cases} \quad (2.20)$$

If $z \in L^6(Q)$ then

$$\frac{\partial z}{\partial t} - \Delta z = z^3 \in L^2(Q) \quad (2.21)$$

so that (we use "fractional" Sobolev spaces, as, for instance, in J.L. LIONS, E. MAGENES [1])

$$z \in L^2(0, T; H^{3/2}(\Omega)) ; \quad (2.22)$$

then (2.21) gives

$$\frac{\partial z}{\partial t} \in L^2(0, T; H^{-1/2}(\Omega)) \quad (2.23)$$

so that z is (a.e. equal to a function) continuous from $[0, T] \rightarrow H^{1/2}(\Omega)$, and therefore

$$z \in L^\infty(0, T; H^{1/2}(\Omega)). \quad (2.24)$$

The cost function is given by

$$J(v, z) = \frac{1}{6} \|z - z_d\|_{L^6(Q)}^6 + \frac{N}{2} \|v\|_{L^2(\Sigma)}^2. \quad (2.25)$$

We want now to find

$$\begin{cases} \inf J(v, z), \text{ for } v, z \text{ subject to (2.19)(2.20) and} \\ v \in U_{ad} = \text{closed convex subset of } L^2(\Sigma); \end{cases} \quad (2.26)$$

we assume that the set of v, z satisfying these constraints is not empty. This problem admits a solution $\{u, y\}$, not necessarily unique. ■

In order to obtain an optimality system (and a Lagrange multiplier) we shall assume (for technical reasons explained below) that

$$\Omega \in \mathbb{R}^2, \quad (2.27)$$

and that

$$z_d \in L^{10}(0, T; L^6(\Omega))^{(1)} \quad (2.28)$$

We have then the existence of p such that

$$p \in L^\infty(0, T; L^2(\Omega)) \cap L^2(0, T; H^1(\Omega)) \quad (2.29)$$

(1) It would be sufficient to assume that $z_d \in L^6(Q)$ and such that there exists $\beta > 0$ such that $z_d \in L^{10}(0, T; L^{5+\beta}(\Omega))$.

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y - y^3 = 0, & -\frac{\partial p}{\partial t} - \Delta p - 3y^2 p = (y - z_d)^5 \text{ in } Q, \\ \frac{\partial y}{\partial \nu} = u, & \frac{\partial p}{\partial \nu} = 0 \text{ on } \Sigma, \\ y(x, 0) = 0, & p(x, T) = 0 \text{ in } \Omega, \end{cases} \quad (2.30)$$

$$\int_{\Sigma} (p + Nu)(v - u) \, d\Sigma \geq 0 \quad \forall v \in U_{ad}, \quad u \in U_{ad}. \quad (2.31)$$

Let us sketch the proof. We start from the penalized problem :

$$J_{\varepsilon}(v, z) = \frac{1}{6} \|z - z_d\|_{L^6(Q)}^6 + \frac{N}{2} \|v\|_{L^2(\Sigma)}^2 + \frac{1}{2\varepsilon} \left\| \frac{\partial z}{\partial t} - \Delta z - z^3 \right\|_{L^2(Q)}^2 \quad (2.32)$$

where $v \in L^2(\Sigma)$, $z \in L^6(Q)$, $\frac{\partial z}{\partial t} - \Delta z \in L^2(Q)$,

$$z(x, 0) = 0 \text{ and } \frac{\partial z}{\partial \nu} = v \text{ on } \Sigma.$$

Then z satisfies (2.22)(2.23)(2.24).

The problem

$$\inf J_{\varepsilon}(v, z), \quad v \in U_{ad} \text{ and } v, z \text{ as above} \quad (2.33)$$

admits (at least) a solution $u_{\varepsilon}, y_{\varepsilon}$. As $\varepsilon \rightarrow 0$ we have

$$u_{\varepsilon}, y_{\varepsilon} \text{ remain in a bounded set of } L^2(\Sigma) \times L^6(Q), \quad (2.34)$$

$$\frac{1}{\sqrt{\varepsilon}} \left(\frac{\partial y_{\varepsilon}}{\partial t} - \Delta y_{\varepsilon} - y_{\varepsilon}^3 \right) \text{ remains in a bounded set of } L^2(Q). \quad (2.35)$$

We introduce

$$p_{\varepsilon} = -\frac{1}{\varepsilon} \left(\frac{\partial y_{\varepsilon}}{\partial t} - \Delta y_{\varepsilon} - y_{\varepsilon}^3 \right). \quad (2.36)$$

We verify that

$$\begin{cases} -\frac{\partial p_{\varepsilon}}{\partial t} - \Delta p_{\varepsilon} - 3y_{\varepsilon}^2 p_{\varepsilon} = (y_{\varepsilon} - z_d)^5, \\ \frac{\partial p_{\varepsilon}}{\partial \nu} = 0 \text{ on } \Sigma, \\ p_{\varepsilon}(x, T) = 0 \text{ in } \Omega \end{cases} \quad (2.37)$$

and

$$(p_{\varepsilon} + Nu_{\varepsilon}, v - u_{\varepsilon})_{L^2(\Sigma)} \geq 0 \quad \forall v \in U_{ad}, \quad u_{\varepsilon} \in U_{ad}. \quad (2.38)$$

It follows from (2.35)(2.34) that

$$\frac{\partial y_{\varepsilon}}{\partial t} - \Delta y_{\varepsilon} = f_{\varepsilon}, \quad f_{\varepsilon} \text{ bounded in } L^2(Q), \quad (2.39)$$

and we have

$$\frac{\partial y_{\varepsilon}}{\partial \nu} = u_{\varepsilon} \text{ on } \Sigma, \quad y_{\varepsilon}(x, 0) = 0 \text{ in } \Omega. \quad (2.40)$$

Therefore (cf. (2.22)(2.23)(2.24))

$$y_\varepsilon \text{ remains in a bounded set of } L^2(0,T;H^{3/2}(\Omega)) \cap L^\infty(0,T;H^{1/2}(\Omega)). \quad (2.41)$$

Therefore, for every θ such that $0 \leq \theta \leq 1$, one has

$$y_\varepsilon \text{ remains in a bounded set of } L^{2/\theta}(0,T;H^{1/2+\theta}(\Omega)). \quad (2.42)$$

But - since the space dimension equals 2 - one has

$$H^{1/2+\theta}(\Omega) \subset L^6(\Omega) \text{ if } \theta = \frac{1}{6}. \quad (2.43)$$

Then (2.42) shows that

$$y_\varepsilon \text{ remains in a bounded set of } L^{12}(0,T;L^6(\Omega)). \quad (2.44)$$

We shall verify below that p_ε satisfies

$$p_\varepsilon \text{ remains in a bounded set of } L^\infty(0,T;L^2(\Omega)) \cap L^2(0,T;H^1(\Omega)). \quad (2.45)$$

Then one can pass to the limit. One shows that $u_\varepsilon, y_\varepsilon, p_\varepsilon \rightarrow u, y, p$ in $L^2(\Sigma) \times L^6(Q) \times \times L^2(0,T;H^1(\Omega))$ weakly (and also in weak topologies or weak star topologies corresponding to (2.41)(2.44)(2.45)), where u, y is a solution of (2.26) and where (2.29) (2.30)(2.31) hold true.

In order to verify (2.45) let us change t into $T-t$ and let us set

$$\begin{aligned} \psi(x,t) &= p_\varepsilon(x, T-t), \quad \sqrt{2} y_\varepsilon(x, T-t) = n_\varepsilon(x,t), \\ (y_\varepsilon(x, T-t) - z_d(x, T-t))^5 &= h_\varepsilon. \end{aligned}$$

We have

$$\begin{cases} \frac{\partial \psi}{\partial t} - \Delta \psi - n_\varepsilon^2 \psi = h_\varepsilon, \\ \frac{\partial \psi}{\partial \nu} = 0 \text{ on } \Sigma, \quad \psi(x,0) = 0 \end{cases} \quad (2.46)$$

$$\text{We set : } |\psi|^2 = \int_{\Omega} \psi^2 dx, \quad \|\psi\|^2 = \int_{\Omega} |\nabla \psi|^2 dx + \int_{\Omega} |\psi|^2 dx.$$

Taking the scalar product of (2.46) with ψ , we obtain

$$\frac{1}{2} \frac{d}{dt} |\psi(t)|^2 + \|\psi(t)\|^2 - |\psi(t)|^2 - (n_\varepsilon^2 \psi, \psi) = (h_\varepsilon, \psi) \quad (2.47)$$

(where $(f, \psi) = \int_{\Omega} f \psi dx$). But

$$|(n_\varepsilon^2 \psi, \psi)| \leq \|n_\varepsilon\|_{L^6(\Omega)}^2 \|\psi\|_{L^2(\Omega)} \|\psi\|_{L^6(\Omega)}$$

so that (and this is valid in dimension $n \leq 3$)

$$\begin{cases} |(n_\varepsilon^2 \psi, \psi)| \leq C \|n_\varepsilon\|_{L^6(\Omega)}^2 |\psi| \|\psi\| \\ \leq \frac{1}{4} \|\psi\|^2 + C \|n_\varepsilon\|_{L^6(\Omega)}^4 |\psi|^2 \end{cases} \quad (2.48)$$

(where the C 's denote various constants).

We have also

$$|(h_\varepsilon, \psi)| \leq \|h_\varepsilon\|_{L^{6/5}(\Omega)} \|\psi\|_{L^6(\Omega)} \leq C \|h_\varepsilon\|_{L^{6/5}(\Omega)} \|\psi\|$$

so that

$$|(h_\varepsilon, \psi)| \leq \frac{1}{4} \|\psi\|^2 + C \|h_\varepsilon\|_{L^{6/5}(\Omega)}^2. \quad (2.49)$$

But

$$\left\{ \begin{aligned} \|h_\varepsilon(t)\|_{L^{6/5}(\Omega)}^2 &\leq C(\|y_\varepsilon(T-t)\|_{L^6(\Omega)}^{10} + \|z_d(T-t)\|_{L^6(\Omega)}^{10}) = \\ &= k_\varepsilon(t), \quad k_\varepsilon \in L^1(0,T), \quad \|k_\varepsilon\|_{L^1(0,T)} \leq C. \end{aligned} \right. \quad (2.50)$$

Therefore (2.47)(2.48)(2.49) give

$$|\psi(t)|^2 + \int_0^t \|\psi(\sigma)\|^2 d\sigma \leq \int_0^t (1+C\|h_\varepsilon\|_{L^6(\Omega)}^4) |\psi|^2 d\sigma + \int_0^t k_\varepsilon(\sigma) d\sigma, \quad (2.51)$$

and since in particular $\|h_\varepsilon\|_{L^6(\Omega)}^4$ remains in a bounded set of $L^1(0,T)$, it follows from (2.51) and Gronwall's inequality that ψ remains in a bounded set of $L^\infty(0,T; L^2(\Omega)) \cap L^2(0,T; H^1(\Omega))$ i.e. (2.45). ■

2.4. Various remarks.

Remark 2.3 : By estimates analogous to those of Section 2.3, one sees that the result of Section 2.2 is valid with U_{ad} not necessarily satisfying (2.16) if $\Omega \subset \mathbb{R}^2$. ■

Remark 2.4 : Let us consider the equation

$$\left\{ \begin{aligned} \frac{\partial y_\varepsilon}{\partial t} - \Delta y_\varepsilon - y_\varepsilon^3 + \varepsilon y_\varepsilon^5 &= v \text{ in } Q, \\ \frac{\partial y_\varepsilon}{\partial \nu} &= 0 \text{ on } \Sigma, \quad y_\varepsilon(x,0) = 0 \text{ in } \Omega. \end{aligned} \right. \quad (2.52)$$

For $\varepsilon > 0$ this problem admits a unique solution $y_\varepsilon(v)$. One can then define

$$J_\varepsilon(v) = \frac{1}{6} \|y_\varepsilon(v) - z_d\|_{L^6(Q)}^6 + \frac{N}{2} \|v\|_{L^2(Q)}^2. \quad (2.53)$$

Then one can verify that :

$$\lim_{\varepsilon \rightarrow 0} (\inf_{v \in U_{ad}} J_\varepsilon(v)) = \inf J(v, z), \quad v, z \text{ subject to (2.1)(2.2)(2.15)}. \quad (2.54)$$

Remark 2.5 : One can obtain similar results for optimality systems connected with stationary problems such as

$$\left\{ \begin{aligned} -\Delta z - z^3 &= v \text{ in } \Omega, \\ \frac{\partial z}{\partial \nu} &= 0 \text{ on } \Gamma \end{aligned} \right. \quad (2.55)$$

and

$$J(v, z) = \frac{1}{6} \|z - z_d\|_{L^6(\Omega)}^6 + \frac{N}{2} \|v\|_{L^2(\Omega)}^2, \quad z_d \in L^6(\Omega), \quad N > 0. \quad \blacksquare$$

BIBLIOGRAPHY

- A.V. FOURSIKOV [1] On some problems of control. Doklady Akad. Nauk, 1980, pp. 1066-1070.
- J.P. KERNEVEZ, J.L. LIONS and D. THOMAS [1] To appear.
- J.L. LIONS [1] Sur le contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles. Paris, Dunod-Gauthier-Villars, 1968 (English translation by S.K. MITTER, Springer, 1971).
- [2] Some aspects of the optimal control of distributed parameter systems. Reg. Conf. Ser. in Applied Math., SIAM, 6, 1972.
- [3] Remarks on the theory of optimal control of distributed systems, in Control Theory of Systems Governed by Partial Differential Equations, ed. by A.K. AZIZ, J.W. WINGATE, M.J. BALAS, Acad. Press, pp. 1-103.
- [4] Function spaces and optimal control of distributed systems. U.F.R. J. Lecture Notes, Rio de Janeiro, 1980.
- [5] Some methods in the Mathematical Analysis of systems and their control. Science Pub. Company, BEIJING, 1981.
- J.L. LIONS and E. MAGENES [1] Problèmes aux limites non homogènes et applications, Vol. 1,2. Dunod, Paris, 1968. English translation by P. KENNETH, Springer Verlag, 1972.
- P. RIVERA [1] To appear
- L. SCHWARTZ [1] Théorie des noyaux. Proc. Int. Congress of Math. 1 (1950), pp. 220-230.
- J. SIMON [1] To appear.

COMPACTNESS AND TOPOLOGICAL METHODS FOR SOME NONLINEAR VARIATIONAL PROBLEMS OF MATHEMATICAL PHYSICS

Pierre L. LIONS
Laboratoire d'Analyse Numérique
Université P. et M. Curie
4, place Jussieu, 75230 Paris Cedex 05
FRANCE

A general method to solve some nonlinear variational problems of Mathematical Physics is illustrated on two examples : the so-called Hartree and Choquard equations. This method is based upon the use of, first, critical point theory and, second, symmetries in order to gain compactness.

INTRODUCTION

We want to present here some general methods to solve some nonlinear variational problems of Mathematical Physics. In general terms, we emphasize the use of critical point theory in order to prove existence and multiplicity results. But, since critical point theory needs some form of compactness (Palais-Smale condition for example) and since in many problems of Mathematical Physics the problem is set in an unbounded domain (like \mathbb{R}^N for example) compactness may be lacking : to avoid this difficulty we will show below that using the symmetries of the problem, we may obtain some form of compactness.

These general (and vague) comments will be illustrated here on two problems : in section I below, we consider the general Hartree equation (and related questions) that is we look for solutions u of :

$$\begin{cases} -\Delta u - \frac{z}{|x|} u + \lambda u + \left(\int_{\mathbb{R}^3} u^2(x-y) \frac{1}{|y|} dy \right) u = 0 \\ u \in H^1(\mathbb{R}^3) \text{ (1)}, u \neq 0 \end{cases}$$

(of course we want non-trivial solutions : $u \neq 0$, since $u \equiv 0$ is a solution of the equation). Here $\lambda \in \mathbb{R}$, $z > 0$ (z is the so-called total charge). As it will be indicated below in section I, we can treat as well general Coulomb potentials

$\left(\sum_i \frac{z_i}{|x - x_i|} \right)$ with $x_i \in \mathbb{R}^3$, $z_i > 0$ instead of $\frac{z}{|x|}$), systems, Hartree-Fock equations and related questions. In section I below, we give general results (that we

believe are optimal) concerning existence of multiple solutions : those results are obtained by a simple use of critical point theory.

Section II is devoted to the study of the so-called Choquard equation, that is we look for solutions u of

$$\begin{cases} -\Delta u + \lambda u - \left(\int_{\mathbb{R}^3} u^2(x-y) \frac{1}{|y|} dy \right) u = 0 & \text{in } \mathbb{R}^3 \\ u \in H^1(\mathbb{R}^3), \quad u \neq 0 ; \end{cases} \quad (2)$$

where $\lambda \in \mathbb{R}$. In section II, we give optimal existence and multiplicity results which are obtained using, in particular, critical point theory. At this point, let us make a remark on (2) : if u is a solution of (2) clearly $u_n(x) = u(x + n\xi)$ (with $n \geq 1$, $|\xi| = 1$) is also a solution of (2), but u_n converges weakly (for example in L^2 or in H^1) to 0 and this convergence cannot be strong (since for example the L^2 norm of u_n is equal to the L^2 norm of u). In other words we have a sequence of solutions (with the same norms in every, say, Sobolev space) that is not compact (in any Sobolev space) : this prevents the direct application of critical point theory (lack of compactness).

As it will be seen in section II, in order to be able to use critical point theory we will use the symmetries of problem (2) - that is of course the spherical symmetry - to obtain some form of compactness.

Finally in section III, we give briefly a list of other problems that can be treated by similar methods (including Nonlinear Scalar Field equations, Rotating Stars, Thomas-Fermi Problems, Vortex Rings, ...). We also present some nonexistence results showing that the symmetries of (in particular) problem (2) are not only essential for the proof of existence results but directly for existence purposes.

Let us finally mention that the results presented in section I are taken from P. L. Lions [26], while those of section III are from P. L. Lions [27]. Finally the solution of all the problems listed in section III can be found in H. Berestycki and P. L. Lions [8],[9],[10], P.L. Lions [28], M.J. Esteban and P. L. Lions [20],[21] ; and a systematic account of all these problems is given in P. L. Lions [29].

I. THE HARTREE EQUATION AND RELATED QUESTIONS

Let us recall the problem : find solutions u of

$$\begin{cases} -\Delta u - z \frac{u}{|x|} + \lambda u + (u^2 * \frac{1}{|x|})u = 0 & \text{in } \mathbb{R}^3 \\ u \in H^1(\mathbb{R}^3), \quad u \neq 0 \end{cases} \quad (1)$$

this problem is some kind of model problem for general Hartree or Hartree-Fock equations (see, for example, E. H. Lieb and B. Simon [25] for the relevance of such problems to Physics) : these equations were introduced by D. Hartree [23] (see also J. C. Slater [36]) as an approximation method to describe the Coulomb Hamiltonian of electrons interacting with static nuclei.

REMARK I.1 : We could treat as well general Hartree or Hartree-Fock systems (see P. L. Lions [26],[29] for more details) : in particular we could replace the Coulomb potential $-\frac{z}{|x|}$ by very general potentials $V(x)$ including for example the general Coulomb potential : $V(x) = -\sum_i \frac{z_i}{|x-x_i|}$ ($x_i \in \mathbb{R}^3$, $z_i \in \mathbb{R}$, $z_i > 0$).

We will indicate below how our main results modify in this case (see P. L. Lions [26],[29] for more details). Finally we could consider the general Thomas-Fermi-Von Weizäcker equations (see R. Benguria, H. Brezis and E.H. Lieb [7] for the introduction of these equations) :

$$\begin{cases} -\Delta u - \sum_i \frac{z_i}{|x-x_i|} u + \beta(u) + \lambda u + (u^2 * \frac{1}{|x|})u = 0 & \text{in } \mathbb{R}^3 \\ u \in H^1(\mathbb{R}^3), \quad u \neq 0 ; \end{cases} \quad (3)$$

where $\lambda \in \mathbb{R}$, $z_i > 0$, $x_i \in \mathbb{R}^3$ and β is some increasing function satisfying $\beta(0) = \beta'(0) = 0$. The results are totally similar to those concerning (1). (2).

REMARK I. 2 : An important role will be played by the eigenvalues of the linearized problem (around 0) namely by the real (s) λ such that there exists $u \neq 0$ solution of :

$$-\Delta u - z \frac{u}{|x|} + \lambda u = 0, \quad u \in H^1(\mathbb{R}^3). \quad (4)$$

It is well-known (see for example E.C.Titchmarsh [40]) that this occurs if and only if $\lambda = \lambda_k = \frac{z^2}{4k^2}$ ($k \geq 1$) and for such $\lambda = \lambda_k$ there exists only one v_k (up to a multiplicative constant) solution of (4).

We may now state our main result :

THEOREM I.1 :

- i) If $\lambda < 0$ or if $\lambda > \lambda_1$, there exists no solution of (1) .
- ii) If $\lambda_{k+1} \leq \lambda < \lambda_k$ ($k \geq 1$), there exist at least k distinct pairs $(\pm u_j)_{1 \leq j \leq k}$ of solutions of (1) such that :

$$u_1(x) > 0 \text{ in } \mathbb{R}^3, \quad S(u_1) = \min_{v \in H^1} S(v) \quad (5)$$

$$S(u_1) < S(u_2) \leq \dots \leq S(u_k) < 0 \quad (6)$$

where $S(v) = \int_{\mathbb{R}^3} \frac{1}{2} |Dv|^2 - \frac{z}{2|x|} v^2 + \frac{\lambda}{2} v^2 dx + \frac{1}{4} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} v^2(x) v^2(y) |x-y|^{-1} dx dy$
for all v in $H^1(\mathbb{R}^3)$.

- iii) If $\lambda = 0$, there exists a sequence of distinct pairs $(\pm u_k)_{k \geq 1}$ of solutions of (1) such that (5) and (6') hold :

$$S(u_1) < S(u_2) < \dots < S(u_k) \rightarrow 0 \text{ as } k \rightarrow +\infty \quad (6')$$

In addition we have : $Du_k \xrightarrow[k]{L^2} 0, \quad u_k \xrightarrow[k]{L^6} 0$.

REMARK I.3 : Let us first explain why S is well defined on $H^1(\mathbb{R}^3)$:
indeed if $v \in H^1(\mathbb{R}^3)$, $v \in L^p(\mathbb{R}^3)$ ($2 \leq p \leq 6$) (Sobolev embeddings)
therefore by classical convolution inequalities : $v^2 * \frac{1}{|x|} \in L^p(\mathbb{R}^3)$

$$(3 < p \leq +\infty) \text{ and } \iint_{\mathbb{R}^3 \times \mathbb{R}^3} v^2(x) v^2(y) |x-y|^{-1} dx dy < +\infty .$$

It is then very easy to prove that $S \in C^1(H^1)$.

REMARK I.4 : If we consider a general Coulomb potential :

$$V(x) = - \sum_i z_i |x-x_i|^{-1} \quad (x_i \in \mathbb{R}^3, z_i > 0) ; \text{ then Theorem I.1}$$

still holds with λ_k being the k -th eigenvalue (If $\lambda_{k-1} > \lambda_k = \lambda_{k+1} = \dots$
 $= \lambda_{k+r} > \lambda_{k+r+1}$, one has in fact for $\lambda < \lambda_k$ ($k+r$) distinct pairs of solutions)
of the Schrödinger operator $-\Delta + V$.

REMARK I.5 : Hartree equations have been studied by many authors (see for example M. Reeken [34], N. Bazley and R.Seydel [5], N. Bazley and B. Zwahlen [6],

K. Gustafson and D. Sather [21], C. A. Stuart [38],[39], J. Wolkowisky [41] but all these works either give local results (λ near λ_1) or use heavily the spherical symmetry and reduce the problem to some O.D.E. question - in particular do not yield any information in the case of a general Coulomb potential. The most general known results - in absence of spherical symmetry - were those of E. H. Lieb and B. Simon [25] but were only concerned with the existence of ground state solutions (essentially those satisfying (5)).

We will only sketch the proof of Theorem I-1, the detailed proof may be found in [26].

REMARK I-6: Using a method due to N. Bazley and R. Seyde [5], one can prove that if $0 \leq \lambda < \lambda_1$, the solution of the minimization problem:

$$\min_{v \in H^1} S(v)$$

is unique (up to a change of sign) and is thus u_1 .

In addition it is obvious to deduce that u_1 , as a function of λ , is C^1 for $0 < \lambda \leq \lambda_1$ and that as $\lambda \rightarrow 0$, $u_1(\lambda)$ converges to $u_1(0)$ in norm in the space $\{u \in L^6(\mathbb{R}^3), Du \in L^2(\mathbb{R}^3)\}$ and weakly in $H^1(\mathbb{R}^3)$.

We now illustrate Theorem I-1 by a few "bifurcation diagrams": let us emphasize that these diagrams are more or less formal. First we need to introduce some notations: let X be the Hilbert space of functions u such that $Du \in L^2(\mathbb{R}^3)$, $u \in L^6(\mathbb{R}^3)$ equipped with the scalar product $((u,v)) = \int_{\mathbb{R}^3} Du(x) \cdot Dv(x) dx$. The bifurcation diagram in X looks like Fig. 1. below:

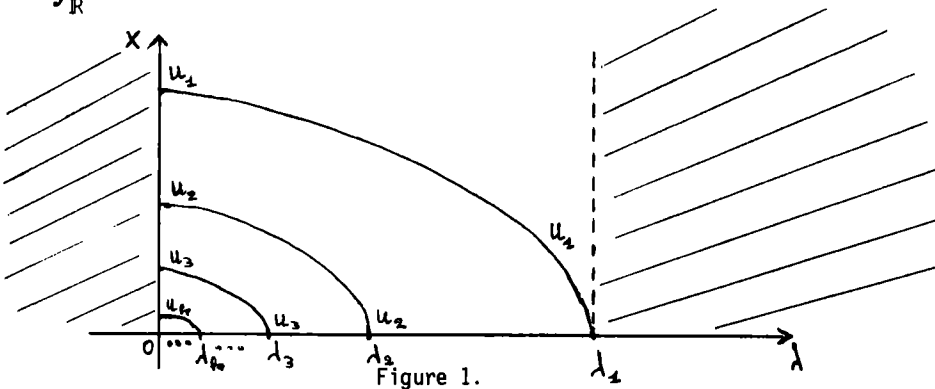


Figure 1.

But now; if we look at the bifurcation diagram in H^1 (or L^2) we find the following Fig. 2:

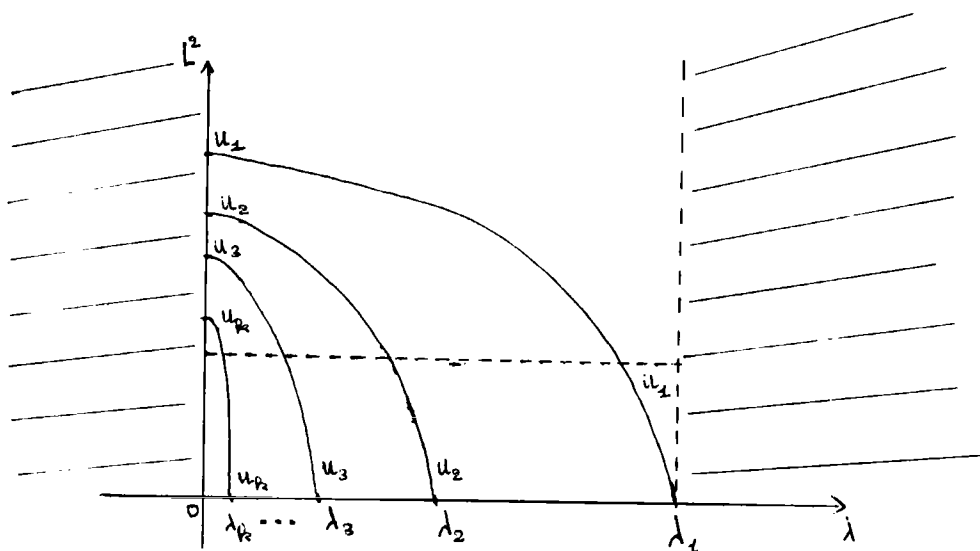


Figure 2

The difference between these two bifurcation diagrams is due to the fact that for $\lambda = 0$, while $u_k \not\rightarrow 0$, $|u_k|_{L^2}$ is bounded away from 0.

We have proved that: $|u_k|_{L^2} > (\frac{2z}{3})^{1/2}$ ($\forall k \geq 1$) and we conjecture that for $\lambda = 0$ we have:

$$|u_k|_{L^2} > z^{1/2} \quad (\forall k \geq 1) \quad \text{and} \quad \lim_{k \rightarrow \infty} |u_k|_{L^2} = z^{1/2}.$$

Finally let us mention that in [25], [7] (by two different methods) it is proved that $|u_1|_{L^2} > z^{1/2}$.

REMARK I-7: Clearly there is bifurcation from each eigenvalue λ_k (which is to be expected in view of the general results of M. G. Grandall and P. H. Rabinowitz [18]). A nice phenomenon is what happens for the branches near $\lambda = 0$: in some sense the branches stop when they hit the continuous spectrum of $-\Delta$ ($\{\lambda \leq 0\}$). More precisely it can be proved (P. L. Lions [29]) that $u_1 (=u_1(\lambda))$ is a C^1 curve in H^1 and that for all $0 < \lambda < \lambda_1$ the linearized operator A_λ around $u_1(\lambda)$ i.e.:

$$H^1 \ni v \mapsto -\Delta v - \frac{z}{|x|} v + \lambda v + (u_1^2 * \frac{1}{|x|})v + 2 \{(u_1 v) * \frac{1}{|x|}\} u_1$$

is coercive and thus is an isomorphism from H^1 onto H^{-1} . But for $\lambda = 0$,

A_0 is only injective and is not onto and this explains why the branch cannot be "extended in the set $\{\lambda < 0\}$ ".

We now turn to the proof of Theorem I-1: we will only give some sketch of the proof, for a detailed proof the reader is referred to P.L. Lions [26].

Sketch of the proof of Theorem I-1 : We will consider only the case when

$0 < \lambda < \lambda_1$ (the non existence part is quite easy and the case when $\lambda = 0$ is quite technical). As it was said before, we will apply critical point theory; and we will just indicate what are the properties satisfied by S here which enable us to apply critical point theory.

Therefore, we claim that $S \in C^1(H^1)$ satisfies :

1) $S(0) = 0$, S is even (and thus $S'(0) = 0$).

2) If $0 < \lambda < \lambda_k$, let $B_k^\varepsilon = \left(\bigoplus_{j=1}^k \mathbb{R} v_j \right) \cap \{ \|u\|_{H^1} = \varepsilon \}$, then

we have on B_k^ε : $S(u) < 0$ for $\varepsilon \in (0, \varepsilon_0]$ (for some $\varepsilon_0 > 0$) .

3) $S(u) \rightarrow +\infty$, as $\|u\|_{H^1} \rightarrow +\infty$

4) S satisfies the Palais-Smale condition :

(6) $\left\{ \begin{array}{l} \text{If } (u_n)_{n \geq 1} \in H^1 \text{ is such that : } S(u_n) \text{ is bounded, } S'(u_n) \xrightarrow{H^{-1}} 0 \\ \text{then there exists a subsequence of } (u_n) \text{ which converges in } H^1 . \end{array} \right.$

Property 1) is obvious; to understand 2) let us just remark that if $\|u\|_{H^1}$ is small then $S(u)$ behaves essentially as its quadratic part which is just the quadratic form associated with $(-\Delta \frac{z}{|x|} + \lambda)$ and thus 2) is clear (recall that v_j is the j -th eigenfunction of $-\Delta - \frac{z}{|x|}$). To explain 3), let us just indicate that if $\lambda > \lambda_{j+1}$ then the main difficulty to check 3) lies with the subspace $\bigoplus_{i=1}^j \mathbb{R} v_i$ of H^1 which is finite dimensional. But on this subspace all norms are equivalent and thus the term $\iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{u^2(x)u^2(y)}{|x-y|} dx dy$ will be predominant at infinity.

Finally 4), which is the fundamental condition in Critical Point Theory, is straightforward to check using 3).

To conclude we just need to apply a result due to Ljusternik and

Schnirelman [31] (for the precise statement, see P.H. Rabinowitz [33], or Krasnoleskii [24]) : conditions 1)-4) imply the existence of at least k distinct pairs of critical points of S , that is solutions of (1) for $\lambda < \lambda_k$.

REMARK I-7 : If one considers the time dependent Hartree equation that is :

$$\begin{cases} i \varphi_t - \Delta \varphi - \frac{z}{|x|} \varphi + (|\varphi|^2 * \frac{1}{|x|}) \varphi = 0 & \text{in } [0, +\infty) \times \mathbb{R}^3 \\ \varphi(x, 0) = \varphi_0(x) \in H^1(\mathbb{R}^3) \end{cases} \quad (7)$$

where φ are complex valued functions.

Then solutions of (1) provide solitary waves for (7) : $\varphi(x, t) = e^{-i\lambda t} u_k(x)$.

In T. Cazenave and P.L. Lions [17], it is proved that the orbit $(e^{-i\lambda t} u_1(x))_{t \geq 0}$ (for $0 \leq \lambda < \lambda_1$) is stable.

II. THE CHOQUARD EQUATION AND RELATED QUESTIONS

We now look for solutions u of

$$\begin{aligned} -\Delta u + \lambda u - (u^2 * \frac{1}{|x|}) u &= 0 & \text{in } \mathbb{R}^3 \\ u &\in H^1(\mathbb{R}^3), \quad u \neq 0. \end{aligned} \quad (2)$$

We will also consider the "normalized" problem : find (λ, u) solution of

$$\begin{cases} -\Delta u + \lambda u - (u^2 * \frac{1}{|x|}) u = 0 & \text{in } \mathbb{R}^3 \\ u \in H^1(\mathbb{R}^3), \quad \lambda \in \mathbb{R}, \quad \|u\|_{L^2}^2 = M \end{cases} \quad (2')$$

where M is given > 0 (for example $M = 1$).

Equations (2) - (2') were introduced by Choquard and the existence of a positive solution of (2') was proved by E.H. Lieb [24]. Problem (2') occurs also in the so-called mean field theory (see M. Donsker and S.R.S. Varadhan [19]). Some other references where such problems are considered are given in P.L. Lions [27].

Our main results concerning (2) or (2') are :

THEOREM II.1 :

- i) If $\lambda \leq 0$, there is no solution of (1)
 ii) If $\lambda > 0$, there exists a sequence of distinct pairs $(\pm u_k)_{k \geq 1}$ of solutions of (2) such that

$$u_1(x) = u_1(|x|) > 0 ; \quad u_1 \text{ is decreasing in } |x| \quad (8)$$

$$0 < S(u_1) \leq S(v) , \text{ for all } v \text{ solution of (2)} \quad (9)$$

$$u_k(x) = u_k(|x|) \quad (\forall k \geq 1) \quad (10)$$

$$0 < S(u_1) < S(u_2) < \dots < S(u_k) \xrightarrow[k \rightarrow +\infty]{} +\infty \quad (11)$$

where $S(v) = \int_{\mathbb{R}^3} \frac{1}{2} |Dv|^2 + \frac{\lambda}{2} v^2 dx - \frac{1}{4} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} v^2(x) v^2(y) |x-y|^{-1} dx dy \quad (\forall v \in H^1).$

THEOREM II.2 :

There exists a sequence $(\lambda_k, \pm u_k)$ of distinct solutions of (2') (for any $M > 0$) such that (8) , (10) hold and :

$$J(u_1) = \min\{J(v)/v \in H^1, |v|_{L^2} = M\} < 0 \quad (9')$$

$$J(u_1) < J(u_2) < \dots < J(u_k) \xrightarrow[k \rightarrow +\infty]{} 0 \quad (11')$$

where $J(v) = \int_{\mathbb{R}^3} \frac{1}{2} |Dv|^2 dx - \frac{1}{4} \iint v^2(y) |x-y|^{-1} dx dy \quad (\forall v \in H^1) ;$

and in addition : $\lambda_k > 0 \quad (\forall k \geq 1), \quad \lambda_k \xrightarrow[k \rightarrow +\infty]{} 0, \quad u_k \xrightarrow[k \rightarrow +\infty]{} 0 \text{ in } L^p$

for $2 < p \leq 6, \quad Du_k \xrightarrow[k \rightarrow +\infty]{} 0 \text{ in } L^2.$

REMARK II-1 : Let us remark that in Theorem II-1, S is not bounded from below (and surely not from above) since $S(tv) \xrightarrow[t \rightarrow +\infty]{} -\infty$ if $v \neq 0$.

REMARK II-2 : The existence of u_1 in Theorem II.2. (with $M = 1$) has been proved by E.H. Lieb [24]. In addition it is proved in [24] that any solution v of :

$$J(v) = \min\{J(w)/w \in H^1, |w|_{L^2} = M\}$$

$$|v|_{L^2} = M, v \in H^1$$

is necessarily of the form : $v(x) = \epsilon u_1(x-x_0)$, for some $x_0 \in \mathbb{R}^3$ and where $\epsilon = \pm 1$. The detailed proofs of Theorems II.1, II.2 can be found in P.L. Lions [27].

REMARK II-3 : If one considers more general equations than (2) of the form :
let $N \geq 2$

$$\left\{ \begin{array}{l} -\Delta u + \lambda u - (u^2 * V(|x|))u = 0 \text{ in } \mathbb{R}^N \\ u \in H^1(\mathbb{R}^N), u \neq 0 \end{array} \right. \quad (12)$$

where V for example is some potential satisfying :

$$V(x) = V(|x|) \geq 0 \text{ in } \mathbb{R}^N \quad (13)$$

$$\begin{aligned} V &\in \mathcal{M}_b(\mathbb{R}^N) + L^q(\mathbb{R}^N) \text{ for some } 1 \leq q < +\infty, \text{ if } N \leq 3 \\ V &\in M^p(\mathbb{R}^N) + M^q(\mathbb{R}^N) \text{ (3) for some } \frac{N}{4} < p < q < +\infty, \text{ if } N \geq 4 \end{aligned} \quad (14)$$

then Theorem II-1 still holds (see P.L. Lions [27], in [27] are also given corresponding extensions of (2')).

In particular, let us remark that $V = \delta_0$ in \mathbb{R}^3 (the Dirac mass at 0) satisfies assumptions (13), (14) and we obtain existence of multiple solutions of

$$-\Delta u + \lambda u = u^3 \text{ in } \mathbb{R}^3, \quad u \in H^1(\mathbb{R}^3), u \neq 0;$$

this had already been proved by several authors (M.S. Berger [16], Z. Nehari [32], G.H. Ryder [35], W. Strauss [27]). Also remark that the above equation arises when looking for solitary waves in cubic nonlinear Schrödinger equations arising in optics (study of laser beams).

Before going into the proofs of Theorems II-1, II-2, let us give the bifurcation diagram corresponding to Theorem II-1 (which can be made rigorous) :

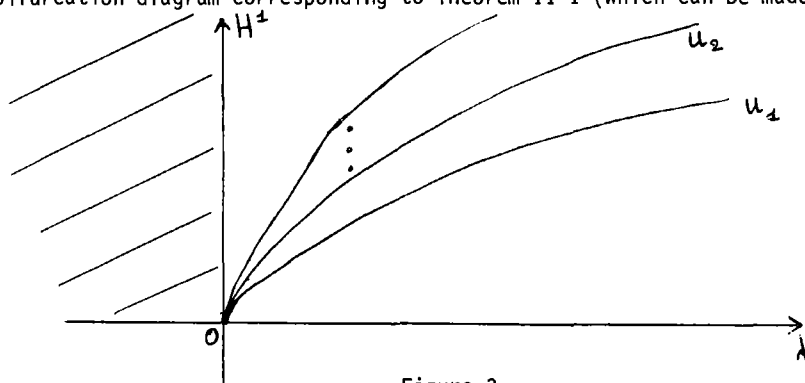


Figure 3.

In particular we see that there is bifurcation from $(0,0)$ of a countable number of branches $\mathcal{C}_k = \{(\lambda, u_k(\lambda))\}$: this phenomena occurs in some nonlinear problems when there is some continuous spectrum (see also H. Berestycki and P.L. Lions [11] for some other example).

We now turn to the proofs of Theorems II.2, II.3 : again we will only sketch it.

Sketch of the proof of Theorem II.1 : We will only consider the case when $\lambda > 0$ (the non existence part is quite easy). Again we will just indicate (without proof) what are the properties satisfied by S which enable us to use critical point theory.

We claim that $S \in C^1(H^1)$ satisfies :

- 1) $S(0) = 0$, S is even (and thus $S'(0) = 0$).
- 2) $\exists \varepsilon, \delta > 0$ such that $S(v) > 0$ if $\|v\|_{H^1} \leq \varepsilon$, $v \neq 0$
and $S(v) \geq \delta > 0$ if $\|v\|_{H^1} = \varepsilon$.
- 3) For any finite dimensional subspace X of H^1 , we have :

$$S(u) \rightarrow -\infty, \text{ if } u \in X, \|u\|_{H^1} \rightarrow +\infty.$$

- 4) S satisfies the Palais-Smale condition in $H_r^1 = \{v \in H^1, v(x) = v(|x|)\}$.

Properties 1) and 2) are clear ; and property 3) is easily deduced from the fact that on X all norms are equivalent. Now 4) is some kind of compactness property that we will explain below.

If one admits these properties, then applying a critical point theorem in the space H_r^1 due to A. Ambrosetti and P.H. Rabinowitz [1], one obtains (essentially) the existence of solutions as in Theorem II.1.

To conclude this sketch of proof, we would like to emphasize the fact that S does not satisfy the Palais-Smale condition in H^1 (in particular because of the difficulty due to translations as mentioned in the Introduction) : but restricting ourselves to the functions having all the symmetries of the problem (here the spherical symmetry), we obtain some compactness. A key lemma in the proof of property 4) is the following Lemma which is a consequence of results due to W. Strauss [37] (see also [8], [29]) :

LEMMA II.1 : The restriction to H_r^1 of the Sobolev embedding from $H^1(\mathbb{R}^3)$ into $L^p(\mathbb{R}^3)$ is compact for $2 < p < \frac{2N}{N-2}$ if $N \geq 3$ and for $2 < p < +\infty$ if $N = 2$.

REMARK II.4 : This lemma clearly shows the role played by the symmetries in order to gain some compactness. Extensions of this Lemma to more general Sobolev spaces

than H^1 is given in P.L. Lions [29], [30].

Sketch of the proof of Theorem II.2 : We just need to apply a general critical point theorem of H. Berestycki and P.L. Lions [12] (see also [9]) : Theorem II.2 is then a consequence of the following properties of J : let

$\mathcal{M} = \{u \in H^1_r, \|u\|_{L^2}^2 = M\}$, we have : $J \in C^1(H^1)$ and

- 1) $J(0) = 0$, J is even,
- 2) J is bounded from below on \mathcal{M} .
- 3) $J|_{\mathcal{M}}$ satisfies Palais-Smale condition.

(Actually $J|_{\mathcal{M}}$ satisfies (P.S.-) condition, and thus something else has to be checked, but we will not enter such technicalities).

Property 1) is obvious, and 3) comes essentially from Lemma II.1. Let us prove 2) : indeed by well-known inequalities :

$$\begin{aligned} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} u^2(x) u^2(y) |x-y|^{-1} dx dy &\leq C \|u\|_{L^{12/5}}^4 \\ &\leq C \|u\|_{L^2}^3 \|u\|_{L^6} \leq C \|u\|_{L^2}^3 \|Du\|_{L^2}. \end{aligned}$$

Therefore, on \mathcal{M} , we have :

$$J(v) \geq \frac{1}{2} \|Dv\|_{L^2}^2 - C M^3 \|Dv\|_{L^2}$$

and this proves that J is bounded from below on \mathcal{M} .

III. RELATED PROBLEMS

III.1 - Symmetries.

In this section, we first mention that if, in view of Lemma II.1, the spherical symmetry gives some compactness, some cylindrical symmetry also gives compactness (see P.L. Lions [28], and also [29], [30]). In particular, we can treat equation (12) when V has some cylindrical symmetry (see [29]).

Now we would like to mention a few non existence results due to M.J. Esteban and P.L. Lions [21] (see also [20]) that seem to indicate that some form of symmetry is needed in order to obtain non trivial solutions of (2) or (12).

For example consider the equations :

$$-\Delta u + \lambda u = |u|^{p-1} u \text{ in } \{x_1 > 0\}, \quad \begin{matrix} u|_{x_1=0} = 0, \\ u(x) \rightarrow 0 \\ |x| \rightarrow +\infty \end{matrix} \quad (15)$$

$$-\sum_{i,j} \frac{\partial}{\partial x_i} (a_{ij}(x) \frac{\partial u}{\partial x_j}) + \lambda u - (u^2 * \frac{1}{|x|}) u = 0, \quad u \in H^1(\mathbb{R}^3) \quad (16)$$

where $\lambda > 0$, $p > 1$, $a_{ij} \in C_b^\infty(\mathbb{R}^3)$ and

$$\exists v \quad \forall x \in \mathbb{R}^N \quad \sum_{i,j} a_{ij}(x) \xi_i \xi_j \geq v |\xi|^2, \quad \forall \xi \in \mathbb{R}^3 \quad (17)$$

$$\left\{ \begin{array}{l} \forall x \in \mathbb{R}^N \quad \sum_{i,j} (\frac{\partial}{\partial x_1} a_{ij}(x)) \xi_i \xi_j \geq 0, \quad \forall \xi \in \mathbb{R}^3 \\ \exists x_0 \in \mathbb{R}^N \quad \sum_{i,j} (\frac{\partial}{\partial x_1} a_{ij}(x_0)) \xi_i \xi_j > 0, \quad \forall \xi \in \mathbb{R}^3 - \{0\} \end{array} \right. \quad (18)$$

Remark that (15) has all symmetries in (x_2, \dots, x_N) : the only broken symmetry is with respect to x_1 . Something similar happens when we assume (18) for the matrix $a_{ij}(x)$.

Then we have the following result :

THEOREM III.1 (M.J. Esteban and P.L. Lions [21]) :

i) The only solution of (15) is : $u \equiv 0$.

ii) The only solution of (16) under assumptions (17)-(18) is : $u \equiv 0$.

In some sense, the symmetries in equations like (12) or (2) keep the solutions from going to infinity and thus vanishing. In the above result "the solutions" slip to infinity and disappear : and we end up with the solution $u \equiv 0$.

III.2 - Other problems :

We conclude by giving a list of other problems which can be treated by similar methods and where the symmetries are essential to get compactness :

1) Nonlinear scalar field equations :

We look for solutions u of

$$\begin{aligned}
 -\Delta u &= g(u) \text{ in } \mathbb{R}^N, \quad u(x) \rightarrow 0, \quad u \not\equiv 0 \\
 |x| &\rightarrow \infty
 \end{aligned} \tag{19}$$

where g is some nonlinearity satisfying : $g(0) = 0$. Of course solutions of (19) are critical points of the following action :

$$S(v) = \int_{\mathbb{R}^N} \frac{1}{2} |Dv|^2 - G(v) dx$$

and the equation (19) is invariant by rotations and translations.

In H. Berestycki and P.L. Lions [8], [9] (see also [13], [14], [15]) this problem is solved under optimal conditions on g by developing new critical point theorems and using the spherical symmetry in order to obtain compactness.

2) Rotating stars :

A model problem is to find $\rho \in L^1(\mathbb{R}^3)$, $\rho \geq 0$, $\int_{\mathbb{R}^3} \rho(x) dx = M$ minimizing, over all $\tilde{\rho} \in L^1(\mathbb{R}^3)$, $\tilde{\rho} \geq 0$, $\int_{\mathbb{R}^3} \tilde{\rho}(x) dx = M$, the functional \mathcal{E} :

$$\mathcal{E}(\tilde{\rho}) = \int_{\mathbb{R}^3} j(\tilde{\rho}) dx + \int_{\mathbb{R}^3} V(r) \tilde{\rho} dx - \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \tilde{\rho}(x) \tilde{\rho}(y) |x-y|^{-1} dx dy \tag{20}$$

where M is given > 0 , j is a nonnegative convex function and V is a given nonnegative bounded function depending only on r (where $x = (r, \theta, z)$ in cylindrical coordinates). This problem is a classical model of equilibrium configurations of axisymmetric rotating fluids.

In P.L. Lions [28] (see also [30]) this problem is solved under optimal conditions on j , V (extending by a different method previous partial results due to J.F.G. Auchmuty and R. Beals [3], see also [2]). The method relies on some new compactness results using the cylindrical symmetry of the problem.

3) General Thomas-Fermi Problems :

A model problem is to find $\rho \in L^1(\mathbb{R}^3)$, $\rho \geq 0$, $\int_{\mathbb{R}^3} \rho(x) dx = M$ minimizing, over all $\tilde{\rho} \in L^1(\mathbb{R}^3)$, $\tilde{\rho} \geq 0$, $\int_{\mathbb{R}^3} \tilde{\rho}(x) dx = M$, the functional \mathcal{E} :

$$\mathcal{E}(\rho) = \int_{\mathbb{R}^3} \frac{3}{5} \rho^{5/3} dx - \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \rho(x) \rho(y) f(x-y) dx dy \tag{21}$$

where M is given > 0 and f is a given spherically symmetric function. Of course this problem has a spherical symmetry.

In P.L. Lions [28], [29], this problem is solved under optimal conditions

using the spherical symmetry to obtain compactness.

4) Vortex rings :

A model problem is to find u satisfying : $\int_{\pi} \frac{1}{r} |Du|^2 dx = \eta$ and minimizing, over all v satisfying $\int_{\pi} \frac{1}{r} |Dv|^2 = \eta$, the functional :

$$J(u) = \int_{\pi} r F((u - wr^2 - k)^+) dx \quad (22)$$

where $\pi = \{(r, z) \in \mathbb{R}^2, r > 0\}$, $w > 0$, $k \geq 0$, $\eta > 0$ are given ; F is a function satisfying : $F(0) = 0$.

Using the symmetry (with respect to z) of the problem, in H. Berestycki and P.L. Lions [10], this problem is solved under optimal conditions.

5) General nonlocal problems :

Here, one wants to find the solutions of :

$$K * u = f(u) \text{ in } \mathbb{R}^N \quad (23)$$

where $K(x) = K(|x|)$, f is a given nonlinearity satisfying : $f(0) = 0$.

This problem is solved in P.L. Lions [29].

Footnotes :

(1) $H^1(\mathbb{R}^3) = \{u \in L^2(\mathbb{R}^3), Du \in L^2(\mathbb{R}^3)\}$

(2) The author would like to thank E.H. Lieb for having pointed out to his attention equation (3).

(3) $M^p(\mathbb{R}^N)$ denotes the usual Marcinkiewicz space.

References

- [1] Ambrosetti, A. and Rabinowitz, P.H., Dual variational methods in critical point theory and applications, J. Funct. Anal. 14 (1973) 349-381.
- [2] Auchmuty, J.F.G., Existence of axisymmetric equilibrium figures, Arch. Rat. Mech. Anal. 65 (1977) 249-261.
- [3] Auchmuty, J.F.G. and Beals, R., Variational solutions of some nonlinear

- free boundary problems, Arch. Rat. Mech. Anal. 43 (1971) 255-271.
- [4] Bader, P., Variational method for the Hartree equation of the Helium atom, Proc. Roy. Soc. Edim. 82 A (1978) 27-39.
 - [5] Bazley, N. and Seydel, R., Existence and bounds for critical energies of the Hartree operator, Chem. Phys. Letters 24 (1974) 128-132.
 - [6] Bazley, N and Zwahlen, B., A branch of positive solutions of nonlinear eigenvalue problems, Manuscripta Math. 2 (1970) 365-374.
 - [7] Benguria R., Brézis, H. and Lieb E.H., The Thomas-Fermi-von Weizsäcker theory of atoms and molecules, to appear in Comm. Maths. Phys.
 - [8] Berestycki, H. and Lions, P.L., Existence of solutions for nonlinear scalar field equations. I. The ground state, to appear in Arch. Rat. Mech. Anal.
 - [9] Berestycki, H. and Lions, P.L., Existence of solutions for nonlinear scalar field equations. II. Existence of infinitely many bound states, to appear in Arch. Rat. Mech. Anal.
 - [10] Berestycki, H. and Lions, P.L., to appear.
 - [11] Berestycki, H. and Lions, P.L., Existence of stationary states of nonlinear scalar field equations. In Bifurcation phenomena in Mathematical physics (Reidel, New-York, 1979).
 - [12] Berestycki, H. and Lions, P.L., Existence d'ondes solitaires dans des problèmes nonlinéaires du type Klein-Gordon, Comptes-Rendus Paris 288 (1979) 395-398.
 - [13] Berestycki, H. and Lions, P.L., Existence of a ground state in nonlinear equations of the type Klein-Gordon. In Variational Inequalities (J. Wiley, New-York, 1979).
 - [14] Berestycki, H. and Lions, P.L., Une méthode locale pour l'existence de solutions positives de problèmes semi-linéaires elliptiques dans \mathbb{R}^N , J. Anal. Math. 38 (1980) 144-187.
 - [15] Berestycki, H., Lions, P.L., and Peletier, L.A., An O.D.E. approach to the existence of positive solutions for semi-linear problems in \mathbb{R}^N , Ind. Univ. Math. J. 30 (1981) 141-157.

- [16] Berger, M.S., On the existence and structure of stationary states for a nonlinear Klein-Gordon equation, *J. Funct. Anal.* 9 (1972) 249-261.
- [17] Cazenave, T., and Lions, P.L., to appear.
- [18] Crandall, M.G., and Rabinowitz, P.H., Bifurcation from simple eigenvalues, *J. Funct. Anal.* 8 (1971) 321-340.
- [19] Donsker, M., and Varadhan, S.R.S., to appear.
- [20] Esteban, M.J., and Lions, P.L., Non existence de solutions non triviales pour des problèmes semilinéaires dans des ouverts non bornés. *Comptes-Rendus Paris* 290 (1980) 1083-1085.
- [21] Esteban, M.J., and Lions, P.L., Existence and non existence results for semilinear elliptic problems in unbounded domains, to appear in *Proc-Boy. Soc. Edim.*
- [22] Gustafson, K., and Sather, D., Branching analysis of the Hartree equations, *Rend. di Mat.* 4 (1971) 723-734.
- [23] Hartree, D., The wave mechanics of an atom with a non-coulomb central field. Part I. Theory and methods, *Proc. Camb. Phil. Soc.* 24 (1928) 89-132.
- [24] Krasnosel'skii, M.A., *Topological methods in the theory of nonlinear integral equations* (Mac Millan, New-York, 1964).
- [25] Lieb, E.H., and Simon, B., The Hartree-Fock theory for Coulomb systems, *Comm. Math. Phys.* 53 (1977) 185-194.
- [26] Lions, P.L., Some remarks on Hartree equation, to appear in *Nonlinear Anal. T.M.A.*
- [27] Lions, P.L., The Choquard equation and related questions, *Nonlinear Anal. T.M.A.* 4 (1980) 1063-1073.
- [28] Lions, P.L., Minimization problems in $L^1(\mathbb{R}^3)$, *J. Funct. Anal.* 41 (1981) 236-275.
- [29] Lions, P.L., Topological and compactness methods for some nonlinear variational problems of Mathematical Physics, to appear.

- [30] Lions, P.L., Minimization problems in $L^1(\mathbb{R}^3)$ and applications to some free boundary problems. In Free boundary problems, Pavia, 1979 (Roma, 1980).
- [31] Ljusternik, L.A., and Schnirelman, L.G., Topological methods in the calculus of variations (Hermann, Paris, 1934).
- [32] Nehari, Z., On a non linear differential equation arising in nuclear physics, Proc. Roy. Irish Acad. 62 (1963) 117-135.
- [33] Rabinowitz, P.H., Variational methods for nonlinear eigenvalue problems. In Eigenvalues of nonlinear problems (Cremonese, Rome, 1974).
- [34] Reeken, M., General theorem on bifurcation and its application to the Hartree equation of the helium atom, J. Math. Phys. 11 (1970) 2505-2512.
- [35] Rvder, G.H., Boundary value problems for a class of nonlinear differential equations, Pacific J. Math. 22 (1967) 477-503.
- [36] Slater, J.C., A note on Hartree's method. Phys. Rev. 35 (1930) 210-211.
- [37] Strauss, W.A., Existence of solitary waves in higher dimensions, Comm. Math. Phys. 55 (1977) 149-162.
- [38] Stuart, C.A., Existence theory for the Hartree equation, Arch. Rat. Mech. Anal. 51 (1973) 60-69.
- [39] Stuart, C.A., An example in nonlinear functional analysis : the Hartree equation, J. Math. Anal. Appl. 49 (1975) 725-733.
- [40] Titchmarsh, E.C., Eigenfunctions expansions, Part II (Oxford Univ. Press, London, 1962).
- [41] Wolkowisky, J., Existence of solutions of the Hartree equations for N electrons. An application of the Schauder-Tychonoff theorem. Ind. Univ. Math. J. 22 (1972) 551-558.

ON YANG-MILLS FIELDS

I. M. Singer

Department of Mathematics
University of California, Berkeley
Berkeley, California 94720
U.S.A.

INTRODUCTION

For this conference on nonlinear problems, I was asked to survey the developments in global nonlinear partial differential equations arising in quantum field theory. I will, in fact, try to give an expository account of the most significant recent development: that of the classification of the self dual Yang-Mills equations of motion. I propose not to be technical because of the diverse backgrounds of the participants. In fact, as a listener, I often get lost early in a talk because the speaker assumes I know more than I do. While he is exposing technical details, I am sadly frustrated struggling with some basic questions whose answers I'm supposed to know.

So I plan to be simple. For the topic I've chosen I'd like to address the following questions:

- (1) What scientific problem led to the nonlinear equation?
- (2) What methods have provided the solution?
- (3) Are these methods applicable to other fields?

Our general experience in mathematics shows that new insights in one area often spread and have other applications. So it is quite useful to understand what the breakthroughs are in a field using nonlinearity; the methods occasionally have other uses.

- (4) How does nonlinearity show itself in the special properties of the solutions?

and finally,

- (5) Do the solutions really shed light on the original scientific question?

Often a nonlinear problem, or any mathematical problem ends up having a life of its own far removed from its place of origin. And one rarely asks if its solutions are relevant to the original motivating question.

OUTLINE

My topic is gauge theories in quantum field theory. I will

- (i) Review the basic setup for gauge fields,
- (ii) Indicate where the nonlinear problem comes from,
- (iii) Describe the problem that has been solved,
- (iv) Point to the new methods that lead to the solutions,
- (v) Exhibit the solutions and some new features they reflect,
- (vi) List some remaining problems,
- (vii) Comment on the usefulness of the solutions in quantum field theory. Rather, I should say, comment on the nonusefulness in quantum field theory to date.

The new features exhibited are (a) compactification; the solutions, which appear to have singularities (sometimes at infinity), really live on compact manifolds and exhibit topological-geometrical properties. Thus a topological discreteness or quantization appears in the form of the Pontryagin charge. Another new feature is (b) complexification; the insight that led to the solutions stems from the complex projective "twistor" program originated by Roger Penrose. Finally, there is (c) excitation, especially of some of us geometers and topologists not accustomed to finding our research applicable to high energy physics.

GAUGE FIELDS

Let me first briefly review gauge fields over Euclidean 4-space R^4 . Euclidean gauge theory on R^4 is easier than the theory over Minkowski $(3+1)$ space. There is a theorem (Osterwalder-Schrader [18]) that says certain constructive field theories in the Euclidean domain with the correct positivity assumptions can be analytically continued back to real time. We are still far away from a constructive gauge theory in 4 dimensions. Still, considerable insight is gained studying quantum field theory in the Euclidean domain.

Let A be a vector potential--or connection, as it is called in mathematics--with values in a simple Lie algebra--say, $su(N)$, the skew-adjoint complex matrices of trace 0. [When I refer to electromagnetism, the Lie algebra is $u(1)$, the pure imaginary numbers.] So A has components A_μ , and the A_μ have values in the Lie algebra $su(N)$:

$$A = A_\mu dx_\mu \quad \text{and} \quad A_\mu = A_\mu^a T_a, \quad (1)$$

where T_a is an orthonormal basis of $su(N)$.

Associated with A is its field or curvature F , a two form with values in $su(N)$:

$$F_{\mu\nu}(A) = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\nu} + [A_\mu, A_\nu]. \quad (2)$$

Note that the bracket $[A_\mu, A_\nu] = A_\mu A_\nu - A_\nu A_\mu$ makes F a nonlinear function of A . All the trouble and all the interest stem from the nonlinear term. It is a feature of the non-Abelian Lie algebra.

The action S is the L^2 norm of the field

$$S(A) = \|F(A)\|^2 = - \sum_{\mu, \nu} \int_{R^4} \text{Tr}(F_{\mu\nu}^2). \quad (3)$$

We are interested in vector potentials with finite action; so let \mathcal{A} be the set of all vector potentials with finite action:

$$\mathcal{A} = \{A; S(A) < \infty\}. \quad (4)$$

Before we can proceed we shall need a few other concepts. First the covariant differential, D_A is the operator

$$(D_A)_\mu = \frac{\partial}{\partial x_\mu} + [A_\mu, \] . \quad (5)$$

Next, a gauge transformation is a function $\phi(x)$ with values in the group G --say, $SU(N)$, the group of $N \times N$ unitary matrices of determinant 1. Let G denote the group of all gauge transformations. G acts on A : a gauge transformation $\phi(x)$ acts on a vector potential A to give the vector potential $\phi \cdot A$ where

$$(\phi \cdot A)_\mu \equiv \phi^{-1} A_\mu \phi + \phi^{-1} \frac{\partial \phi}{\partial x_\mu} ; \quad (6)$$

or, equivalently

$$D_{\phi \cdot A} = \phi^{-1} D_A \phi . \quad (7)$$

It is easy to check that

$$F_{\mu\nu}(\phi \cdot A) = \phi^{-1} F_{\mu\nu}(A) \phi . \quad (8)$$

Since ϕ has values in a compact group--here $SU(N)$ --formula (8) implies that the action is invariant under G :

$$S(\phi \cdot A) = S(A) . \quad (9)$$

Finally, we need some geometry of 4-space. Let $*$ denote duality of two forms in R^4 : $12 \rightarrow 34$, $13 \rightarrow -24$, $14 \rightarrow 23$, and $*^2 = +1$. So $(*F)_{12} = F_{34}$, $(*F)_{13} = -F_{24}$ and $(*F)_{14} = F_{23}$. Also let

$$F^\pm \equiv \frac{F \pm *F}{2} . \quad (10)$$

Duality decomposes the field F into its self dual F^+ and anti-self dual F^- components; these components are the (+1) and (-1) eigenvalues under $*$.

Ordinary two forms--not just two forms with values in $SU(N)$ --split the same way. Two forms are denoted by Λ^2 , the space of skew symmetric 2-tensors. The splitting gives

$$\Lambda^2 = \Lambda_+^2 + \Lambda_-^2 \quad (11)$$

where Λ_\pm^2 are the self-dual and anti-self-dual ordinary forms respectively. Formula (11) gives the decomposition of $\mathfrak{so}(4)$ into two $\mathfrak{so}(3)$'s--or, since the Lie algebras are the same--into two $\mathfrak{su}(2)$'s.

Note that

$$S(A) = ||F(A)||^2 = ||F^+||^2 + ||F^-||^2$$

For later use we define

$$\rho(A) = \frac{||F^+||^2 - ||F^-||^2}{8\pi^2} . \quad (13)$$

Observe that $*$ is conformally invariant in 4 dimensions, so that $S(A)$ is a conformal invariant. Shortly we shall see how the 4-sphere S^4 , the conformal compactification of R^4 , enters.

THE PROBLEM

We can now state the mathematical question: What are the stationary points of $S(A)$? The motivation for the problem stems from the saddle point method of evaluating the Feynman-Kac path integral

$$\langle f(A) \rangle = \int_A DA e^{-S(A)} f(A) / \int_A DA e^{-S(A)} \quad (14)$$

where $f(A)$ is a gauge invariant quantity: $f(\phi \cdot A) = f(A)$.

It is a fundamental problem of quantum field theory to "make sense" of the integral above--one really doesn't know what DA , "integration over all fields" means. In the saddle point method one fixes on the stationary points. Then one writes the action as a quadratic term in A plus a remainder, and uses a perturbative expansion for the remainder term. One emulates quantum electrodynamics (QED) where the method has been spectacularly successful. It is essential, then, to know the stationary solutions.

It is easy to compute the Euler-Lagrange equations for the stationary points:

$$D_A^+ F(A) = 0 , \quad (15)$$

where $+$ indicates the adjoint operator; in components,

$$\frac{\partial F_{\mu\nu}}{\partial x_\mu} + [A_\mu, F_{\mu\nu}] = 0 . \quad (16)$$

We seek, then, the solutions A to (16) with finite action on R^4 .

EXTENSION TO S^4

One of the points I wish to emphasize in this lecture is that consideration of

the problem above leads naturally to the 4-sphere, to topology and fibre bundles; these concepts are not introduced artificially.

To this end I cite a theorem of Karen Uhlenbeck [20].

Theorem: Suppose

- (i) $S(A) < \infty$,
- (ii) A satisfies the equation of motion, i.e., is a stationary point, and
- (iii) A has an isolated (point) singularity.

Then the singularity is removable by a gauge transformation in the following sense: if the singularity is at p , there is a ball (D_p) around p and a gauge transformation ϕ defined on $D_p - p$ so that $\phi \cdot A$ is extendable to all of D_p . [See Fig. 1.]

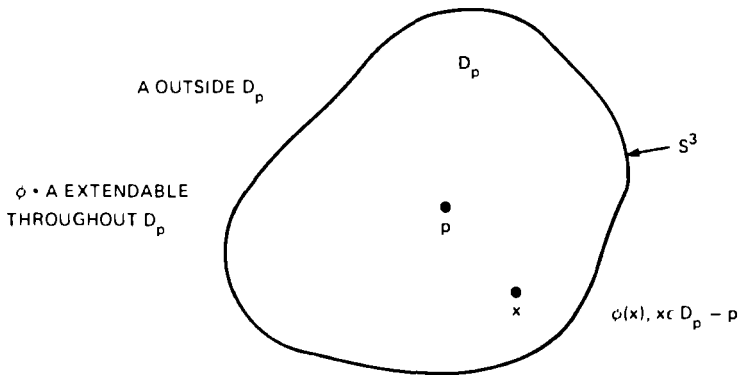


Figure 1: The domain D_p around a point p used in extending the vector potential A to remove a point singularity.

We have split R^4 into two parts: the Ball D_p , and its complement; their common boundary is the 3-sphere S^3 . On D_p , the vector potential $\phi \cdot A$ is nonsingular and there is a mismatch between A and $\phi \cdot A$ on the common boundary S^3 . We use it to build a new object--a fibre bundle.

Before describing the fibre bundle it is worth noting that we can allow the singularity to be at infinity. Because of conformal invariance, infinity is no different than any other point in R^4 ; that is, a conformal transformation brings infinity back to the origin. So a finite action, stationary solution--with no singularities on R^4 , but not defined at infinity--can be considered to have an isolated singular point at ∞ . By a conformal transformation, bring ∞ back to the origin and apply Uhlenbeck's theorem to remove the singularity. Figure 1 is now modified and becomes Fig. 2. The ball D_p is the upper hemisphere D_+ , its complement is the lower hemisphere D_- , and their common boundary is the equator S^3 .

Now we are ready to build a principal bundle over the 4-sphere, and a vector potential associated with it.

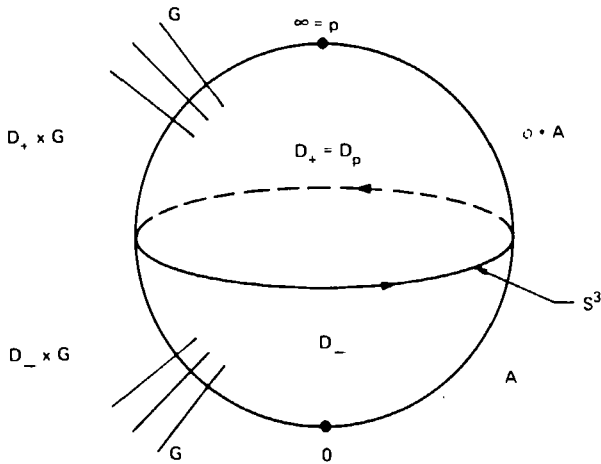


Figure 2: The construction of a fibre bundle over the 4-sphere.

The simplest principal bundle over the 4-sphere is $S^4 \times G$. The trouble is the vector potential A is singular. To remove the singularity, we break A up into $\phi \cdot A$ in the upper hemisphere and A in the lower hemisphere producing a mismatch between A and $\phi \cdot A$ along the equator. We compensate for this mismatch by building a new bundle. On the upper hemisphere it is $D_+ \times G$. We patch or paste along the equator by identifying the set $x \times G$ of $D_+ \times G$ with $x \times G$ of $D_- \times G$ by left multiplication with $\phi(x)$. We get a fibre bundle, that is, a collection of fibres, groups in this case, one for each point of S^4 . But the fibres have been sheared along the equator by the gauge transformation $\phi(x)$, $x \in S^3$. Since the gauge transformation $\phi(x)$ on S^3 with values in G can be topologically non-trivial, i.e., can wind around G , the newly constructed principal bundle P can be complicated. The point of its construction is that it compensates for the mismatch of A and $\phi \cdot A$ on the equator S^3 . We have produced a new vector potential $A \cup_{S^3} \phi \cdot A$ on P without singularities.

Uhlenbeck's theorem says, in short, that any stationary point of the action on R^4 with isolated singularities can be promoted to a nonsingular vector potential on a principal bundle over S^4 . The new vector potential is a stationary point for the action on this principal bundle. One can recapture the original picture on R^4 by stereographic projection.

A DIVERSION

Fibre bundles have become so common in high energy physics (grand unification schemes, symmetry breaking, and dimensional reduction), that perhaps a short digression into their geometry is in order. A few pictures in the simplest cases are worth many formulas, in my view. Consider, then Table I and Fig. 3.

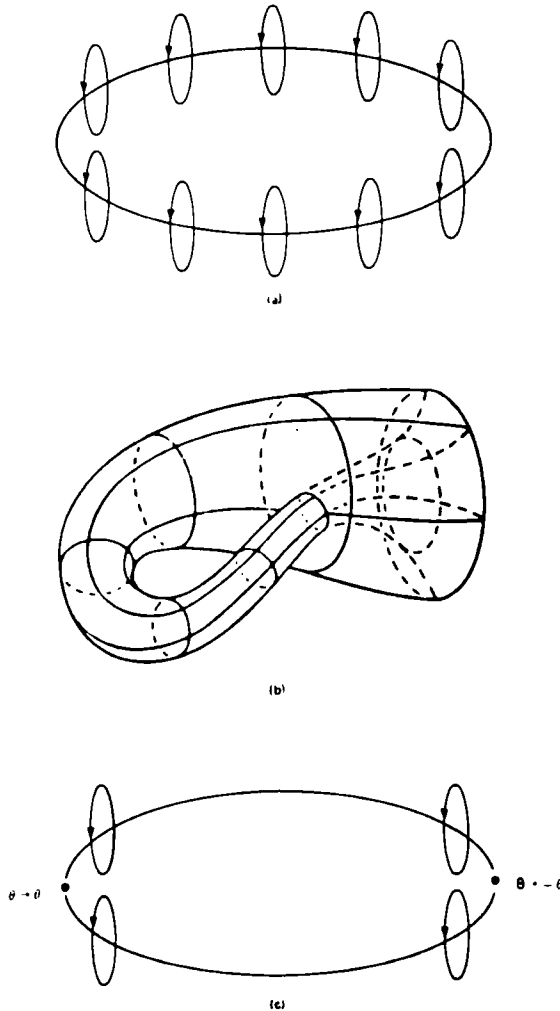


Figure 3: Fibre bundles over the circle S^1 : (a) the torus; (b) the Klein bottle; (c) an alternative description of the Fibre bundle leading to a Klein bottle.

TABLE I
FIBRE BUNDLES

- | | |
|--|---|
| 1. Circle bundles over a circle S^1 : | torus, Klein bottle |
| 2. Circle bundles over a 2-sphere S^2 : | The Chern-Gauss-Bonnet formula and Dirac Magnetic Monopoles |
| 3. Principal bundles over a 4-sphere S^4 : | Our previous examples; classification in terms of the Pontryagin charge or second Chern class |

On the circle (S^1) there are two circle bundles, one of which is trivial and one that isn't. You have seen them both. The first is a torus. Take a circle as shown in Fig. 3a, and bring it around the base circle. The second is a Klein bottle (Fig. 3b): when the fibre circle is brought around the base circle to the starting point, the identification is made with the initial fibre by $\theta \leftrightarrow -\theta$ instead of $\theta \leftrightarrow \theta$. A slightly different way of looking at this construction generalizes to higher dimensions. Take the base S^1 and break it up into two pieces, so there is an upper hemisphere (arc) and lower hemisphere. Each hemisphere is an interval and the "equator" consists of two points. The only issue is how to patch the fibre circles at the equator. At one end patch by the identity and at the other end patch by $\theta \leftrightarrow -\theta$ giving the Klein bottle (see Fig. 3c).

Next consider the two sphere (S^2) (Fig. 4), decomposed into the upper hemisphere (D_+) and the bottom hemisphere (D_-). We want to construct a bundle of circles by patching or pasting $D_+ \times S^1$ and $D_- \times S^1$ along the equator. To do so requires a map ϕ from the equator (a circle) into the circle S^1 . We know these maps ϕ are determined topologically by their winding number. Patching by ϕ at x on the equator means identifying the circle from the northern hemisphere with the circle from the southern hemisphere by a rotation through angle $\phi(x)$. The topology of the bundle is completely determined by the winding number of ϕ which has the formula

$$\frac{1}{2\pi i} \int \frac{\phi'}{\phi} . \quad (17)$$

As an example, consider the circle bundle of unit tangent vectors of S^2 (Fig. 4a). The fibre at each point is the circle of unit tangent vectors emanating from that point. What is the winding number for this bundle? We can easily identify all the fibres in the upper hemisphere by adopting a relativistic point of view. An observer at the north pole carries a clock (a unit tangent vector) along with him as he moves along a great circle path (geodesic) to any other point in the northern hemisphere. Since the unit tangent vectors of a geodesic are auto-parallel, the direction ξ is carried into the direction ξ' , determining the identification of the circle labeled C' with the one labelled C (Fig. 4b). Do the same for an observer at the south pole; a little surface theory, tells you the shear is twice the solid angle Ω . Thus, the winding number is two.

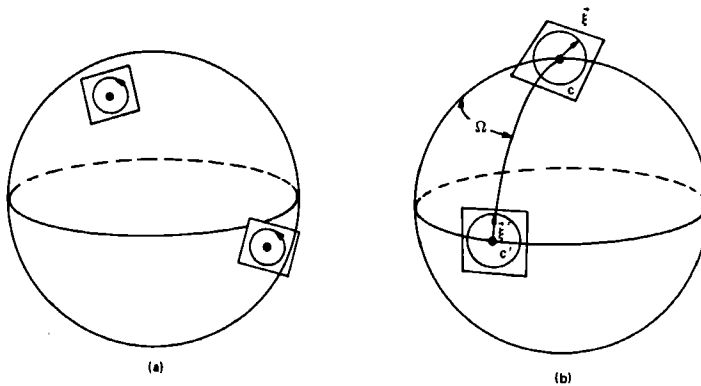


Figure 4: Construction of a fibre bundle on the 2-sphere, S^2 .

Here is another geometric example to help build your intuition for the four dimensional case we'll need later. Think of the 3-sphere, S^3 , as the set of pairs of complex numbers (Z_1, Z_2) subject to

$$|Z_1|^2 + |Z_2|^2 = 1 \quad . \quad (18)$$

Map the 3-sphere to the two sphere by sending a pair (Z_1, Z_2) into Z_1/Z_2 .

$$\begin{aligned} S^3 &= [(Z_1, Z_2) : |Z_1|^2 + |Z_2|^2 = 1] \\ \downarrow S^1 \quad \downarrow \mathbb{C} \quad (e^{i\theta} Z_1, e^{i\theta} Z_2) & \\ S^2 \quad Z_1/Z_2 & \end{aligned} \quad (19)$$

For our present purposes, S^2 equals the complex plane plus ∞ because Z_1/Z_2 can be ∞ if $Z_2 = 0$ and $|Z_1| = 1$. Since $|Z_1|^2 + |Z_2|^2 = 1$, multiplying by $e^{i\theta}$ rotates vectors in S^3 but preserves the ratio: $Z_1/Z_2 = e^{i\theta} Z_1/e^{i\theta} Z_2$. Moreover if $Z_1/Z_2 = W_1/W_2$ and (Z_1/Z_2) is of the same length as (W_1/W_2) , then $Z_j = e^{i\theta} W_j$, $j = 1, 2$. So we have exhibited S^3 as a family of circles over S^2 : a circle bundle where each circle projects into the same ratio or point of S^2 . This bundle has winding number one, though that's not easy to see. It's computable by the Gauss-Bonnet formula which we now discuss.

Let's again think of the two sphere as unit vectors in three space. Dirac found all circle bundles over S^2 , motivated by magnetic monopoles [10]. He started out with Maxwell's equations for an electromagnetic field F . In concise mathematical notation, F satisfies

$$\begin{aligned} dF &= 0 \\ d^*F &= j, \text{ the current} \quad . \end{aligned} \quad (20)$$

Dirac wanted to reverse the two equations above to obtain

$$\begin{aligned} dF &= \rho_M(0), \text{ a magnetic pole at the origin} \\ d^*F &= 0 \quad . \end{aligned} \quad (21)$$

He was interested in the static case, so the equations reduce to three space. Away from the origin, say on a sphere of fixed positive radius, $dF = 0$. On a region that is contractible, one could conclude that $F = dA$. But the sphere is not contractible so $F = dA$ on the entire sphere. In fact, we know that the integral of F over the two sphere has to be the flux of the magnetic field and it is not zero. But the upper hemisphere, D_+ is contractible and we can write

$$F|_{D_+} = dA_+ \text{ on } D_+ \quad . \quad (22)$$

Similarly, on the lower hemisphere, $F|_D = dA_-$. However, A_+ and A_- will not agree along the equator; there is a mismatch. In a little band about the equator $dA_+ = F = dA_-$ so

$$d(A_+ - A_-) = 0 \quad . \quad (23)$$

One can show that

$$A_+ - A_- = \frac{d \log \phi}{i} = \frac{d\phi}{i\phi} \quad \text{on the equator} \quad (24)$$

for some function ϕ that doesn't vanish. [In general $\int (A_+ - A_-)$ around the circle could be any number. But if matter fields are introduced and A_+ and A_- are to be vector potentials, $\int A_+ - A_-$ is an integer and Eq. (24) holds, thus quantizing the magnetic pole at the origin.] The flux will be (up to a universal factor) equal to the winding number of ϕ . That is

$$\begin{aligned} \text{flux} &= c \frac{1}{2\pi} \int_{S^2} F = \frac{c}{2\pi} \int_{D_+} F + \frac{c}{2\pi} \int_{D_-} F = \frac{c}{2\pi} \int_{\partial D_+} A_+ - \frac{c}{2\pi} \int_{\partial D_-} A_- \\ &= \frac{c}{2\pi} \int_{S^1} A_+ - A_- = \frac{c}{2\pi i} \int \frac{d\phi}{\phi} = c \text{ winding no. of } \phi \quad . \end{aligned} \quad (25)$$

The function ϕ can be used to build a circle bundle as I have already described. That is, circles from the top hemisphere are patched with circles from the bottom hemisphere by $\phi/|\phi|$. Now A_+ and A_- fit together to give a vector potential on this circle bundle over S^2 . The circle bundle is determined by the winding number which we can compute in two different ways. First, we can calculate it as the flux; this is the generalized Chern, Gauss-Bonnet theorem, and gives the winding numbers as an integral formula, the averaged curvature field. Second, given the vector potential, by parallel transport along geodesics from the north and south poles (as we did with the tangent circle bundle), we can compute the mismatch at the equator and thus compute the winding number.

Now let's return to the 4-sphere, S^4 , and reconsider Fig. 2. We had two hemispheres, D_+ and D_- , and the equator S^3 . The $SU(N)$ bundles are determined by maps of S^3 into $SU(N)$. When $N = 2$, $SU(2)$ is the same as the unit quaternions, i.e., S^3 . The maps of $S^3 \rightarrow S^3$ are determined up to homotopy by an integer, the winding number k . (This remains true for arbitrary N .) We have already seen that the vector potentials on R^4 with isolated singularities satisfying the equations of motion can be promoted to vector potentials on these bundles over S^4 . What is the corresponding formula for k in terms of an integral formula for the field? The answer is again the Chern-Gauss-Bonnet formula: in terms of F , the self-dual, and \bar{F} , the anti-self-dual part of the field

$$k = \rho(A) = \frac{1}{8\pi^2} (||F^+(A)||^2 - ||F^-(A)||^2) \quad . \quad (26)$$

Finally, let me end the geometric discussion with an example of an $SU(2)$ bundle over S^4 that generalizes (19). We use quaternions instead of complex numbers.

$$\begin{aligned}
 S^7 &= [q_1, q_2), |q_1|^2 + |q_2|^2 = 1] \\
 \downarrow S^3 &\quad \downarrow (qq_1, qq_2) \\
 S^4 &\quad q_2^{-1}q_1
 \end{aligned} \tag{27}$$

Consider S^7 as pairs of quaternions with the sum of absolute values squared = 1. Since q_1 and q_2 are four dimensional, the restriction $|q_1|^2 + |q_2|^2 = 1$ defines a sphere in eight-dimensional space, i.e., S^7 . Map S^7 to S^4 by mapping (q_1, q_2) into $q_2^{-1}q_1$, also a quaternion. S^4 is to be interpreted as quaternions plus infinity, because when $q_2 = 0$, $q_2^{-1}q_1$ equals infinity. As before, if $|q| = 1$, then (qq_1, qq_2) is also on S^7 and $(qq_2)^{-1}qq_1 = q_2^{-1}q_1$. Furthermore, if two points on S^7 give the same ratio, they differ by a unit quaternion ratio. Hence we get the 3-sphere bundle, S^7 over S^4 .

One interesting fact that emerges from our discussion is that the topological quantization is inherent in the problem. We have shown that certain stationary solutions could be promoted to S^4 as vector potentials on bundles over S^4 . A vector potential with finite action may oscillate near infinity, and may not be promotable to S^4 . Nevertheless, the right-hand side of Eq. (26) makes sense. It's a real number and not infinity because it is less than the action. One can ask whether it's always an integer as in the stationary case. The answer is yes, if mild decay conditions are assumed for A as $x \rightarrow \infty$. [Then $A_\mu \sim \phi^{-1} \partial_\mu \phi$ and ϕ , as a function of rays, has a winding number.] So the set of "most" finite action vector potentials splits into a discrete set (quantized by the integer k) called Pontryagin sectors.

THE SOLUTION

I have taken considerable time describing our problem and its promotion to a geometric-topological context. Now I want to talk about solutions. We still do not know all the stationary solutions (up to a gauge transformation). It is known [7] that a local minimum, in a given Pontryagin sector, is a minimum. One can now say a great deal about the minima.

By subtracting Eq. (26) from Eq. (12), we get

$$S(A) = 8\pi^2 k + 2||F^-||^2. \tag{28}$$

We can assume $k > 0$, since a change of orientation changes the sign of k . Thus F equal to 0, i.e., $F = F^+$, assures us a minimum. Such F are called "self-dual" fields. What do these self-dual fields look like? Note that, whereas being stationary leads to a second-order equation, the condition for a self-dual minimum gives a first order equation; it is much easier to deal with.

Theorem.[6] Self-dual solutions exist for $N = 2$ and for all $k \geq 0$.

There's an Ansatz for them: If f is a harmonic function,

$$A_\alpha = \frac{1}{2} \eta_{\alpha\beta} \frac{\partial_\beta f}{f}, \quad (29)$$

is self-dual. The matrix $\eta_{\alpha\beta}$ --I'm using physics notation here-- is just the projection of the two-form $e^\alpha \wedge e^\beta$ onto $\Lambda^2 = \mathfrak{su}(2)$ so that A_α is a Lie algebra-valued form. The specific harmonic functions used are

$$f = 1 + \sum_{j=1}^k \frac{\lambda_j}{|x - x(\rho_j)|^2} \quad (30)$$

and give the "multi-instantion" solutions [13,14]. They have isolated singularities at ρ_j . By Uhlenbeck's theorem, the singularities can be removed; these vector potentials have Pontryagin charge k .

One can count the number of degrees of freedom for the solutions above. One finds:

TABLE 2

k	Number of parameters ($N = 2$)
0	unique solution
1	5 degrees of freedom
2	13 degrees of freedom
$k > 2$	$5k + 4$

One can also count the number of degrees of freedom by linearizing the differential equation for self duality. The linear equation is a standard elliptic one and the index theorem counts the number of degrees of freedom. For $SU(2)$, it gives $8k - 3$. For $SU(N)$ we have the

Theorem: [3] The self-dual solutions are a manifold of $\dim 4Nk - N^2 + 1$.

For $N = 2$ we find $8k - 3 > 5k + 4$ when $k \geq 3$, so there must be additional solutions for $k \geq 3$. What do they look like? Surprisingly, the answer comes from algebraic geometry. And it is the insight of the Penrose twistor program that translates the self-dual problem to a holomorphic one [4,19,21]. Briefly, what happens is this. The sphere S^7 in (28) is acted upon by the circle $e^{i\theta}$ (see Eq. (31)). CP^3 denotes complex projective space, the complex lines in complex 4-space; that is $[(z_1, z_2, z_3, z_4) \sim (\lambda z_1, \lambda z_2, \lambda z_3, \lambda z_4); \lambda \in \mathbb{C} \text{ and } \lambda \neq 0]$. This space is a fibre bundle over S^4 with fibre S^2 . The map is given by $(z_1, z_2, z_3, z_4) \mapsto z_1 + z_2 j / z_3 + z_4 j$.

$$\begin{array}{ccc} CP^3[z_1, z_2, z_3, z_4 \sim (\lambda z_1, \lambda z_2, \lambda z_3, \lambda z_4)] & & (31) \\ \downarrow & \downarrow & \\ S^4 & (z_1 + z_2 j / z_3 + z_4 j) & \end{array}$$

The $SU(N)$ bundle P over S^4 pulls up to a bundle over CP^3 , and the fibre can be complexified to $SL(N, \mathbb{C})$ [see Eq. (32)].

$$\begin{array}{ccc} CP^3 & \xleftarrow{SL(N, \mathbb{C})} & P^C \\ \downarrow & & \downarrow \\ S^4 & \xleftarrow{SU(N)} & P \end{array} \quad (32)$$

That is, the $SU(N)$ fibre over a point p of S^4 is lifted to all the S^2 points in CP^3 over p . The group $SU(N)$ is enlarged to the complex special linear group $SL(N, \mathbb{C})$, constructing a new principal bundle P^C over CP^3 . The vector potential A can be promoted to a vector potential A^C on P^C . Self duality for A translates immediately into the Nirenberg-Newlander conditions [17] for making P^C , with connection A^C , into a holomorphic bundle. That is the insight provided by the Penrose Twister Program.

Put more concretely, the ratios z_1/z_2 , z_2/z_4 , z_3/z_4 are coordinates for CP^3 (when $z_4 \neq 0$). Two ratios would suffice to coordinatize S^4 locally; but all three complex coordinates are needed to translate self dual into holomorphic (even though the third coordinate is redundant and acts somewhat like gauge freedom).

There is a similar effect in two-dimensional abelian Yang-Mills/Higgs theory [15]. There the action $S(A, \phi)$ depends on an abelian ($G = U(1)$) vector potential A and a complex scalar field ϕ :

$$S(A, \phi) = ||F_A||^2 + ||D_A \phi||^2 + \lambda/4 ||(1 - |\phi|^2)||^2 \quad (33)$$

Critical points are complicated. The equations of motion are:

$$D_A^* F = \text{Imag. } \bar{\phi} D_A \phi, \quad (D_A D_A^* + D_A^* D_A) \phi = \frac{1 - |\phi|^2}{2} \phi \quad (34)$$

Note that to have finite action, $|\phi| \rightarrow 1$ as $r \rightarrow \infty$ and $||D_A \phi|| \rightarrow 0$ as $r \rightarrow \infty$. So $A_\mu = \phi^{-1} \partial_\mu \phi + O(1/r^2)$. Again the finite action configurations fall into classes determined by the winding number $k = \frac{1}{2\pi i} \int_g \phi' / \phi = \frac{1}{2\pi i} \oint A_\mu = \frac{1}{2\pi i} \int_{R^2} F_{\mu\nu} \epsilon_{\mu\nu} d^2x$.

For $\lambda = 1$, the minima are much easier to determine. The equations are:

$$D_A \phi = i^* D_A d, \quad *F = \frac{1 - |\phi|^2}{2} \quad (35)$$

Equation (35) says that $\partial_{\bar{z}} \cdot i[(A_1 + iA_2)/2] (\phi) = 0$. That is $e^{-w} \phi$ is holomorphic where $\bar{\partial} w = (A_1 + iA_2)/2$. The solutions ϕ are of the form $\phi = h(z) \prod_l (z - z_l)^{n_l}$ with the winding number $k = \sum n_l$.

The reduction to a holomorphic problem is again predictable from the twister program, which also leads to new non-Abelian magnetic monopoles [15].

Once the self-dual problem is translated to a holomorphic one on CP^3 , it becomes a problem in algebraic geometry. The algebraic geometers have been studying this matter extensively; that is, studying the classification of holomorphic bundles over CP^3 [5,12]. Those that come from the self-dual vector potentials on the four spheres can be classified. We describe them in two ways [2]. First, the abstract description for the complete set of solutions is this: consider linear maps, $T(z)$, $z \in C^4$, from C^k to C^{2k+2} ; these are $2k+2$ by k matrices, depending linearly on z . Assume the second space C^{2k+2} has a symplectic structure given by a skew-adjoint matrix, J . Also assume that $T(z)$ satisfies these conditions: (1) It's of maximal rank so it maps C^k injectively into C^{2k+2} ; that is, for any vector v in $T(z)(C^k)$, $\langle Jv, v \rangle = 0$. Then

$$E(z) = \frac{T(z)(C^k)^1}{T(z)(C^k)} \quad (36)$$

is two dimensional. For $T(z)(C^k)$ is isotropic and lies inside the J -orthogonal complement which is of dimension $k+2$. The two-dimensional space $E(z)$ is independent of the scale and this construction assigns to every line in C^4 or every point of CP^3 a two-dimensional space. The $SU(2)$ bundle is the set of orthonormal frames in each $E(z)$.

Theorem. Every $SU(2)$ -bundle with self-dual vector potential can be represented this way.

For a thorough discussion of this theorem and its generalization to $SU(N)$ see [1].

Despite the abstract construction, the self-dual vector potentials can be realized quite concretely, as illustrated by the Ansatz I describe next. First let me remark that any vector potential can be written as

$$A_\mu = T^+ \frac{\partial T}{\partial x_\mu}, \quad (37)$$

where T is an L by N matrix function with $T^+ T = I_N$ (see Narasiman and Varadan [16]). These A_μ will be self-dual if T satisfies certain conditions. First, for the k th Pontryagin sector, L is $N+2k$. Next, note that the condition $T^+ T = I_N$ means the columns of T are N orthogonal vectors in C^{N+2k} . Choose $2k$ vectors in C^{N+2k} orthogonal to the columns of T . Denote this $N+2k \times 2k$ matrix by Δ . Hence, $T \Delta = 0$.

Self duality for A_μ is given by the following equations in Δ :

$$\Delta(x) = a + bx, \quad \text{and} \quad (38a)$$

$$\Delta^+ \Delta \text{ commutes with multiplication by quaternions.} \quad (38b)$$

In Eq. (38a), the point x in R^4 stands for quaternionic multiplication. That is, $x = x_0 + x_1 i + x_2 j + x_3 k$ with i, j, k represented by the usual 2×2 matrices. Also, $a = (a_1, a_2)$ and $b = (b_1, b_2)$ with a_i, b_i constant $N+2k \times k$ matrices. Equation (38b) says that $\Delta^+ \Delta$ is a $k \times k$ matrix $\square I_2$ [8].

It is easy to check that vector potentials satisfying this Ansatz are self dual. One can also check that the number of degrees of freedom is $8k - 3$ for $N = 2$. The algebraic geometry assures us that all solutions are gauge equivalent to the exhibited solutions. It is amusing to note that if one could prove directly that the space of all solutions is connected, then the detour through geometry would be unnecessary. For the exhibited solutions form a complete manifold of the proper dimension.

OPEN PROBLEMS

We close with a list of open problems.

- (a) Are there irreducible stationary solutions which are not self (anti) dual? See [11,23] for a twistor description of the equations of motion.
- (b) What are the complex solutions of the Yang-Mills equations of motion, i.e., solutions for \mathbb{C}^4 with $G = \text{SL}(N, \mathbb{C})$?
- (c) Is the space of self-dual solutions connected for $k \geq 3$?
- (d) If no decay assumptions are made, does a vector potential A with finite action have $k = 1/8\pi^2 (||F||^2 - ||F^*||^2)$ an integer?
- (e) What are the applications to quantum gauge theories?

So far, the existence of solutions of different topological type, i.e., the existence of distinct Pontryagin sectors, describes the vacuum structure of the theory (a'la tunnelling). The specific forms for the solution as displayed by Eqs. (37) and (38) imply a nice closed form [9] for the Green's function of the Dirac operator ϕ_A coupled to A :

$$G(x,y) = T^+(y) T(x)/||x-y||^2 \quad (39)$$

and this form has been useful in computing determinants in the weak coupling limit.

But attempts to interpret or compute Feynman integrals in Quantum Chromodynamics using the self-dual solutions have not been successful to date. The "dilute gas approximation" doesn't work and, as far as I can tell, the configurations of self- and anti-dual solutions that dominate the integral have not been found.

REFERENCES

- [1] Atiyah, M. F., Geometry of Yang-Mills Fields, Fermi Lectures 1979 (Pisa).
- [2] Atiyah, M. F., Drinfeld, V. G., Hitchin, N. J., and Manin, Yu. I., Construction of instantons, Phys. Lett. 65A (1978), pp. 185-187.
- [3] Atiyah, M. F., Hitchin, N. J., and Singer, I. M., Self-duality in four dimensional Riemannian geometry, Proc. Roy. Soc. London, A362 (1978), pp. 425-461.
- [4] Atiyah, M. R. and Ward, R. S., Instantons and algebraic geometry, Commun. Math. Phys. 55 (1977), pp. 117-124.

- [5] Barth, W., Some properties of stable rank-2 vector bundles on P_n , Math. Ann. 226 (1977), pp. 125-150.
- [6] Belavin, A., Polyakov, A., Schwartz, A., and Tyupkin, Y., Pseudoparticle solutions of the Yang-Mills equations, Phys. Lett. 59B (1975), pp. 85-87.
- [7] Bourguignon, J.-P., Lawson, H. B., and Simons, J., Stability and gap phenomena for Yang-Mills fields, Proc. Nat. Acad. Sci. U.S.A. 76 (1979), pp. 1550-1553.
- [8] Christ, N., Stanton, N., and Weinberg, E. J., General self-dual Yang-Mills solutions, Columbia University preprint, 1978.
- [9] Corrigan, E., Fairlie, D. B., Goddard, P., and Templeton, S., A Green's function for the general self-dual gauge field, preprint (E.N.S., Paris).
- [10] Dirac, P.A.M., Proc. Roy. Soc. A133, 60 (1931).
- [11] Green, P. S., Isenberg, J., and Yasskin, P., Non-self-dual gauge fields, Phys. Lett. 78B (1978), pp. 462-464.
- [12] Horrocks, G., Vector bundles on the punctured spectrum of a local ring, Proc. London Math. Soc. 14 (1964), pp. 684-713.
- [13] Jackiw, R., Nohl, C., and Rebbi, C., Classical and semi-classical solutions to Yang-Mills theory, Proceedings 1977 Banff School, Plenum Press.
- [14] Jackiw, R., Nohl, C., and Rebbi, C., Conformal properties of pseudo-particle configurations, Phys. Rev. 150 (1977), pp. 1642-1646.
- [15] Jaffe, A. and Taubes, C., Vortices and Monopoles, Birkhauser (1980) Boston.
- [16] Narasimhan, M.S. and Ramanan, S., Existence of universal connections, Amer. J. Math. 83 (1961), pp. 563-572.
- [17] Newlander, A. and Nirenberg, L., Complex analytic coordinates in almost complex manifolds, Ann. of Math. 65 (1957), pp. 391-404.
- [18] Osterwalder, K. and Schrader, R., Axioms for Euclidean Green's functions I, II, Comm. Math. Phys. 31 (1973); 42 (1975).
- [19] Penrose, R., The Twistor programme, Rep. Mathematical Phys. 12 (1977), pp. 65-76.
- [20] Uhlenbeck, K., Removable singularities in Yang-Mills fields, Bull. Amer. Math. Soc. i (1979), pp. 579-581.
- [21] Ward, R.S., On self-dual gauge fields, Phys. Lett. 61A (1977), pp. 81-82.
- [22] Ward, R.S., A Yang-Mills-Higgs monopole of charge 2, Commun. Math. Phys. 79 (1981), pp. 317-325.
- [23] Witten, E., An interpretation of classical Yang-Mills theory, Phys. Lett. 77B (1978), pp. 394-398.

GAUGE THEORIES FOR SOLITON PROBLEMS

D.H. Sattinger^{*}

Department of Mathematics
University of Minnesota
Minneapolis, Minn.
U.S.A.

1. Lax equations and isospectral deformations.

Gardner, Greene, Kruskal, and Miura [4] showed that the eigenvalues of the Schrodinger operator $L = \partial_x^2 + \frac{1}{6}u$ are invariant if the potential $u = u(x, t)$ evolves according to the KdV equation $u_t + u_{xxx} + uu_x = 0$. P. Lax [5] gave this observation the following general formulation: Suppose $L(u)$ is a linear operator depending on the function u such that the family of operators $\{L(u(t))\}$ is unitarily equivalent when u evolves according to a nonlinear evolution equation $\mathcal{E}(u, u_x, u_{xx}, u_t, u_{xt}, \dots) = 0$. This means that there is a one parameter family of unitary operators $U(t)$ such that $U^{-1}(t)L(u(t))U(t) = L(u(0))$. The unitary operators must satisfy a differential equation of the type $U_t = BU$, where $B^* = -B$. Differentiating $U^{-1}LU$ with respect to t we obtain the equation

$$\partial_t L = [B, L] \quad (1)$$

In the context of soliton theories this equation has sometimes been called the Lax equation, and the operators B and L a Lax pair for the evolution equation $\mathcal{E} = 0$. In the case of the KdV equation Lax gave for B the operator $B = -4(\partial_x^3 + \frac{1}{4}u\partial_x + \frac{1}{8}u_x)$. Then $\partial_t L = \frac{1}{6}u_t$ (multiplication) and one can check that the commutator $[B, L]$ is the multiplication operator $-\frac{1}{6}(u_{xxx} + uu_x)$. Thus $L_t = [B, L]$ is an operator equation which is satisfied whenever u is a solution of the KdV equation.

The KdV equation can be solved explicitly by the method of inverse scattering

* This research was supported in part by NSF Grant #MCS 78-00415.

applied to the scattering problem $(L - \lambda)\varphi = 0$. Some of the other striking properties of the KdV equation are its Hamiltonian structure, an infinite number of conservation laws, an associated invariant (isospectral) scattering problem, and multi-soliton solutions.

Since the first investigations of the KdV equation it has become apparent that the KdV equation is not alone in possessing these properties. Other evolution equations with many of the same properties are the nonlinear Schrodinger equation $i\psi_t + \psi_{xx} + |\psi|^2\psi = 0$, which was shown to share similar properties by Zakharov and Shabat [10]; and the Sine-Gordon equation, which was investigated by Ablowitz, Kaup, Newell, and Segur [1].

These three equations, as well as others, can all be obtained in the context of a gauge theory with gauge group $SL(2, \mathbb{R})$. The gauge theory of these equations will be discussed in §4; and in §5 we discuss isospectral problems from the general point of view of gauge theories on fiber bundles. In [8] we derived a generalization of the nonlinear Schrodinger equations for the gauge group $sl(3, \mathbb{C})$. In §6 we derive multi-component analogues of the sine/sinh-Gordon equations for a general semi-simple Lie algebra. We illustrate these equations explicitly for the case of algebras of rank two, namely A_2, B_2 and G_2 . Similar equations have been obtained by Fordy and Gibbons [3] and Mikhailov *et. al.* [6].

In the gauge theory point of view the Lax equation (1) arises as the condition of zero curvature of connection $\{D_x, D_t\}$ on a principal bundle: thus (1) takes the form $[D_x, D_t] = 0$. The role of the scattering operator L is then played by D_x and that of the infinitesimal generator of the unitary transformations by D_t . We begin in §§2,3 with a discussion of Wahlquist and Estabrook's notion of prolongation algebras and show how the isospectral problems and the connection may be introduced from their point of view.

I wish to thank Leon Green and Peter Olver for many useful discussions concerning the work in this paper.

2. Prolongation Albegras [9]

The first step in the Wahlquist-Estabrook program is to replace the evolution equation by an ideal of 2-forms closed under exterior differentiation. This procedure is best illustrated by an example, say the sine-Gordon equation

$u_{xt} = \sin u$. We write this as a first order system $u_x = p$, $p_t = \sin u$: and then, in x, t, u, p space we introduce the two forms

$$\alpha_1 = du \wedge dt - p \, dx \wedge dt, \quad \alpha_2 = dp \wedge dx + \sin u \, dx \wedge dt.$$

The four dimensional space with coordinate x, t, u, p is regarded as a bundle:

$\mathbb{R}^2 = \{u, p\}$ over $\mathbb{R}^2 = \{x, t\}$. A section of this bundle is a graph of the form $(x, t, u(x, t), p(x, t))$. When α_1 and α_2 are restricted to such a section they take the form

$$\tilde{\alpha}_1 = (u_x - p) dx \wedge dt, \quad \tilde{\alpha}_2 = (\sin u - p_t) dx \wedge dt.$$

Thus, α_1 and α_2 vanish on solution manifolds of the sine-Gordon equations.

According to Cartan's theory, the equations $\tilde{\alpha}_1 = \tilde{\alpha}_2 = 0$ are equivalent to the partial differential equation $u_{xt} = \sin u$ iff the ideal generated by $\{\alpha_1, \alpha_2\}$ is closed under exterior differentiation. A simple computation shows that

$$d\alpha_1 = dt \wedge \alpha_2, \quad d\alpha_2 = \cos u \, dx \wedge \alpha_1,$$

so indeed this ideal is closed.

The Cartan ideal (the ideal generated by the two forms; in the case above, that generated by α_1 and α_2) is now prolonged by introducing additional variables y_1, \dots, y_n ; one-forms

$$\omega^k = dy^k + F^k \, dx + G^k \, dt; \quad (2)$$

and requiring that $\{\omega^1, \dots, \omega^n, \alpha_1, \alpha_2, \dots\}$ be closed under exterior differentiation. The functions F^k and G^k are assumed to depend only on the dependent variables u, p (as the case may be) and the $y^1 \dots y^n$. In the case at hand,

$$d\omega^k = dF^k \wedge dx + dG^k \wedge dt$$

$$\begin{aligned}
&= F_u^k du \wedge dx + F_p^k dp \wedge dx + \frac{\partial F^k}{\partial y^j} dy^j \wedge dx \\
&= G_u^k du \wedge dt + G_p^k dp \wedge dt + \frac{\partial G^k}{\partial y^j} dy^j \wedge dt .
\end{aligned}$$

From (2) $dy^j \wedge dx = \omega^j \wedge dx - G^j dt \wedge dx$ and $dy^j \wedge dt = d\omega^j \wedge dt - F^j dx \wedge dt$.

Also, from the definition of α_1 and α_2 , $dp \wedge dx = \alpha_2 - \sin u dx \wedge dt$ and $du \wedge dt = \alpha_1 + p dx \wedge dt$. The prolongation equations are obtained by substituting these identities in the expression above for $d\omega^k$ and then setting $d\omega^k \equiv 0$ modulo the ideal generated by α_1, α_2 , and ω^k . In the case of the Sine-Gordon equation we obtain

$$\begin{aligned}
G_u &= 0 & G_p &= 0 \\
[G, F] + pG_u - \sin u F_p &= 0
\end{aligned}$$

Prolongation equations for other well known evolution equations have also been derived. Wahlquist and Estabrook have derived the prolongation equations for the KdV and nonlinear Schrodinger equations. The problem now is to solve the prolongation equations for vector fields F and G . This was originally done by Wahlquist and Estabrook for the KdV equation [9]. The case of the Sine-Gordon equation is much simpler (see [8]).

3. Isospectral Problems

Let us turn now to the problem of finding associated isospectral problems for a given evolution equation. We suppose that vector fields F and G have been found which satisfy the prolongation equations and are linear in the y -variables:

$F = y^j F_j^k \frac{\partial}{\partial y^k}$ and $G = y^j G_j^k \frac{\partial}{\partial y^k}$. This gives us an isomorphism between the Lie algebra of matrices and of vector fields. A Lax pair is given in terms of the matrices $F = F_j^k$ and $G = G_j^k$ by $L = \partial_x - G$ and $B = -G$. The associated isospectral scattering problem is given by $Ly = 0$ or $\frac{\partial y^j}{\partial x} - F_k^j y^k = 0$.

Let us verify this for the sine-Gordon equations. From the prolongation equations we have

$$L_t = F_u u_t + F_p p_t = F_p p_t$$

and

$$[B, L] = -[G, L] = -[G, \partial_x - F]$$

$$\begin{aligned}
 &= \partial_x G - [F, G] \\
 &= p G_u - (p G_u - \sin u F_p) = \sin u F_p
 \end{aligned}$$

The sine-Gordon equation thus comes from the Lax equation $L_t = [B, L]$. We can get explicit realizations of the isospectral problems by taking specific representations of the Lie algebra so (3). For example, by choosing the representation of so (3) given by the Pauli spin matrices

$$\sigma_1 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \sigma_2 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

one is led to the scattering problem introduced by Ablowitz et. al. [1]. (See [8]).

4. Connections on Fiber Bundles; $\mathfrak{sl}(2, \mathbb{R})$

The Lax equations can be interpreted from the point of view of gauge theories, as follows. Consider the case of the sine-Gordon equation, and let ∂_x, ∂_t be the total derivatives on the jet bundle x, t, u, p

$$\begin{aligned}
 \partial_x &= \frac{\partial}{\partial x} + p \frac{\partial}{\partial u} + p_x \frac{\partial}{\partial p} + \dots \\
 \partial_t &= \frac{\partial}{\partial t} + u \frac{\partial}{\partial u} + p \frac{\partial}{\partial p} + \dots
 \end{aligned}$$

Putting $D_x = \partial_x - F$, $D_t = \partial_t - G$, we have, from equations (3)

$$\begin{aligned}
 [D_x, D_t] &= [\partial_x - F, \partial_t - G] \\
 &= \partial_t F - \partial_x G + [F, G] \\
 &= F_p p_t - G_u p + p G_u - \sin u F_p \\
 &= (p_t - \sin u) F_p.
 \end{aligned}$$

Thus the curvature $[D_x, D_t]$ of the connection vanishes on an integral manifold of the sine-Gordon equation.

It is useful to compare this formalism with the gauge theories of force fields.

For example, in electromagnetic theory one introduces the connection

$$D_\mu = \partial_\mu + e A_\mu$$

where $\mu = 0, 1, 2, 3$; the ∂_μ are the derivatives $\frac{\partial}{\partial x^\mu}$, and the A_μ are the electromagnetic 4-potentials. In electromagnetic theory the A_μ are scalars, for the gauge group of electromagnetic theory is the abelian group $U(1)$; but in non-abelian gauge theories the A_μ take their values in a Lie algebra. The

commutator $[D_\mu, D_\nu] = F_{\mu\nu}$ gives the field strengths. In electromagnetic theory $F_{\mu\nu} = e(\partial_\mu A_\nu - \partial_\nu A_\mu)$. Thus the condition of zero curvature $[D_\mu, D_\nu] = 0$ means zero external field. The Lax equations may be interpreted as the vanishing of the curvature of a connection, or, as we shall see below, as an integrability condition.

To see how to proceed let us take $sl(2, \mathbb{R})$ as gauge group; $sl(2, \mathbb{R})$ gauge theories lead to the sine-Gordon, KdV and nonlinear Schrodinger equations. As basis for $sl(2, \mathbb{R})$ we may take

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

These matrices have the commutation relations $[\sigma_1, \sigma_2] = 2\sigma_3$, $[\sigma_1, \sigma_3] = -2\sigma_2$, $[\sigma_2, \sigma_3] = \sigma_1$. Form the connection

$$D_x = \partial_x + i\zeta\sigma_1 - q\sigma_2 - r\sigma_3$$

$$D_t = \partial_t - A\sigma_1 - B\sigma_2 - C\sigma_3.$$

Setting the curvature $[D_x, D_t] = 0$ we are led to the system of equations

$$\begin{aligned} -\partial_x A + qC - rB &= 0 \\ q_t - \partial_x B - 2(qA + i\zeta B) &= 0 \\ r_t - \partial_x C + 2(rA + i\zeta C) &= 0 \end{aligned} \quad (5)$$

which were obtained by Ablowitz et. al. [2]. The coefficients A, B, C depend only on q, r and their derivatives. Ablowitz et. al. construct solutions of (5) by expanding

$$A = \sum_{n=0}^N \zeta^n A_n, \quad B = \sum_{n=0}^N \zeta^n B_n, \quad C = \sum_{n=0}^N \zeta^n C_n;$$

these series may be substituted into (5) and the coefficients A_n, B_n, C_n solved for recursively. The evolution equations are obtained as the coefficients of the $\zeta^0 = 1$ terms.

Examples: (1) Nonlinear Schrodinger equation

$$\begin{aligned} r &= \bar{\Psi}, \quad q = \Psi \\ A &= -2i\zeta^2 - i|\Psi|^2 \\ B &= 2\zeta\Psi + i\Psi_x \\ C &= 2\zeta\bar{\Psi} - i\bar{\Psi}_x \end{aligned}$$

$$[D_x, D_t] = (i\Psi_t + \Psi_{xx} + |\Psi|^2 \Psi) \sigma_2 + (-i\Psi_t + \Psi_{xx} + |\Psi|^2 \bar{\Psi}) \sigma_3 .$$

$$(ii) \quad KdV$$

$$r = -1, \quad q = u$$

$$A = -4i\zeta^3 + 2i\zeta u - u_x$$

$$B = 4\zeta^2 u + 2i\zeta u_x - 2u^2 - u_{xx}$$

$$C = -4\zeta^2 + 2u$$

$$[D_x, D_t] = (u_t + 6uu_x + u_{xxx}) \sigma_2 .$$

The sine-Gordon equation is obtained by putting $A = \frac{a}{\zeta}$, $B = \frac{b}{\zeta}$, $C = \frac{c}{\zeta}$:

$$D_x = \partial_x + i\zeta \sigma_1 + u_x \left(\frac{\sigma_2 - \sigma_3}{2} \right)$$

$$D_t = \partial_t - \frac{1}{4i\zeta} \cos u \sigma_1 - \frac{1}{4i\zeta} \sin u (\sigma_2 + \sigma_3)$$

$$[D_x, D_t] = (u_{xt} - \sin u) \left(\frac{\sigma_2 - \sigma_3}{2} \right) .$$

In this formalism

$$\text{Lax equations} \quad \longleftrightarrow \quad [D_x, D_t] = 0$$

$$\text{isospectral problem} \quad \longleftrightarrow \quad D_x(\zeta)\varphi = 0$$

$$\text{Unitary group} \quad \longleftrightarrow \quad D_t\varphi = 0 .$$

That is, as q and r evolve according to the nonlinear evolution equation the eigenfunctions (solutions of $D_x(\zeta)\varphi = 0$) must evolve according to the differential equation $D_t\varphi = 0$.

5. Connections on Fiber Bundles; general.

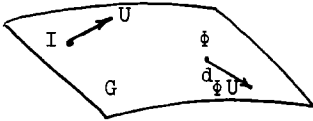
We consider a principal bundle of a linear (matrix) Lie group G over \mathbb{R}^2 ; and let $\{D_x, D_t\}$ be a connection on this principal bundle. Let Φ be the fundamental solution matrix of $D_x(\zeta)\Phi = 0$. (We assume D_x contains the parameter ζ as one of the coefficients, as in the case of $sl(2, \mathbb{R})$ above).

Proposition. If $[D_x, D_t] = 0$ there exists a single-valued solution $\Phi(x, t, \zeta)$ of the equations $D_x(\zeta)\Phi = 0$, $D_t\Phi = 0$. For each point x, t Φ takes its values in the Lie group G . Φ is thus a section of the principal bundle of G over \mathbb{R}^2 .

Proof. The equation $D_x\Phi = 0$ is of the form

$$\partial_x \Phi = U\Phi$$

where U lies in the Lie algebra \mathfrak{g} . This can be regarded as a differential equation on the Lie group G .



The mapping $A \rightarrow A\Phi$ on G takes the identity matrix I to Φ . Since $A\Phi$ is matrix

multiplication, the Jacobian of this mapping is $d_\Phi U = U\Phi$. Hence the differential equation can be interpreted as

$$\frac{d\Phi}{dx} = d_\Phi U$$

where $d_\Phi U$ is a tangent vector to G at Φ . Hence, if $\Phi(x_0, t_0)$ lies in the group, the entire trajectory $\{\Phi(x, t_0)\}$ lies in G . The same argument applies to the equation $D_t \Phi = 0$. Since D_x and D_t commute, the flows generated by these operators commute; and $\Phi(x, t)$ depends only on its value at a base point (x_0, t_0) , not on the path.

Now suppose that the coefficient functions in the D_x component tend to zero as $|x| \rightarrow \infty$; and that at $x = \pm \infty$ $D_x \sim \partial_x - U_\pm$ where U_\pm is an element of the Lie algebra \mathfrak{g} . Then $D_x \Phi = 0$ implies

$$\Phi \sim e^{U_+ x} A_+ \quad \text{as } x \rightarrow +\infty$$

where the A_\pm are constant matrices. Choose Φ so that

$$\Phi \sim e^{U_+ x} \quad \text{as } x \rightarrow -\infty$$

Then there is a matrix S such that $\Phi \sim e^{U_+ x} S(t, \zeta)$ as $x \rightarrow +\infty$. We suppose that $D_t \sim \partial_t - V_+$ as $|x| \rightarrow \infty$.

Proposition 2. If $[U_+, V_+] = 0$ the time evolution of the "scattering data" $S(t, \zeta)$ is given by

$$S_t - V_+ S = 0.$$

Proof. We may assume the asymptotic behavior $\Phi \sim e^{U_+ x} S(t, \zeta)$ is uniform in t and that this relation may be differentiated in time. Since $D_t \Phi = 0$ we have

$$D_t e^{U_+ x} S(t, \zeta) = e^{U_+ x} S_t - V_+ e^{U_+ x} S = e^{U_+ x} (S_t - V_+ S) \sim 0 \quad \text{as } x \rightarrow \infty.$$

Since $S_t - V_+ S$ is independent of x and $e^{U_+ x}$ is nonsingular, $S_t - V_+ S = 0$

It is proposition 2 which accounts in part for the success of the inverse scattering method. One can solve for the evolution of the scattering data S with no knowledge of the solutions of the nonlinear evolution equations. The reader may check that the condition $[U_+, V_+] = 0$ is satisfied in each of the cases for $sl(2, R)$ described above.

6. Multi-component sine-Gordon equations.

Blowitz et. al. [2] arrived at the sine-Gordon and sinh-Gordon equations by assuming $A(\zeta, q, r) = \frac{a(q, r)}{\zeta}$, etc. in (5). These equations are of the form $u_{xt} = f(u)$, where f is a sum of two exponentials. In this section we extend their method to a general semi-simple Lie algebra g to obtain, as an analogue, a system of equations $u_{xt}^i = f^i(u^1 \dots u^k)$ where each f^i is a sum of exponentials. Some such systems have been discussed by Fordy and Gibbons [3], and A.V. Mikhailov et. al. [6], but by other methods.

The number k of variables in our system is equal to the rank of the semi-simple algebra g (i.e. the dimension of the Cartan subalgebra). The number of independent exponentials is determined in a more complicated way from the roots of g .

In the case $sl(3, C)$, for example, we will have two independent variables and the nonlinear terms will be sums of three exponentials. The algebra $sl(3, C)$ is A_2 . We construct the equations for A_2 , B_2 , and the exceptional Lie algebra G_2 .

We begin our construction with the ansatz

$$D_x = \partial_x - \zeta \sigma_1 - u^i \sigma_i$$

$$D_t = \partial_t - \frac{a^j}{\zeta} \sigma_j$$

where i runs from 2 to n . Setting $[D_x, D_t] = 0$ we get

$$-\lambda_x a^j \sigma_j + \sum_{i=2}^n u^i a^j [\sigma_1, \sigma_j] = 0$$

$$\partial_t u^i \sigma_1 + a^j [\sigma_1, \sigma_j] = 0$$

In both equations j is summed over 1 to n and i over 2 to n .

The first can be written

$$\partial_x A = \sum_{i \neq 1} [u^i \sigma_i, A]$$

where $A = a^j \sigma_j$. We try to integrate this system by setting $u^i = \frac{\partial \varphi^i}{\partial x}$ and

$$A = A(\varphi^2, \dots, \varphi^n). \text{ We get}$$

$$\frac{\partial A}{\partial \varphi^i} = [\sigma_i, A]$$

The necessary conditions for integrability of this system are

$$[\sigma_j, [\sigma_1, A]] = [\sigma_1, [\sigma_j, A]]$$

for $i, j = 2, \dots, n$. By the Jacobi identity we may replace this by $[[\sigma_1, \sigma_j], A] = 0$

The easiest way to satisfy this condition is to choose the σ_1 to lie in an abelian subalgebra. We are thus led to try

$$D_x = \partial_x - u^i h_i - \zeta \sigma \quad D_t = \partial_t - \frac{b^j}{\zeta} \sigma_j$$

where $\{h_i\}$ are elements in the Cartan subalgebra and σ, σ_j are root vectors.

(There is no loss in generality (proof omitted) by omitting components along the Cartan subalgebra in D_t).

The condition $[D_x, D_t] = 0$ then leads to

$$\partial_x b^j = u^i \alpha_j(h_i) b^j \quad (\text{no sum over } j) \quad (6)$$

$$\partial_t u^i h_i + b^j [\sigma, \sigma_j] = 0.$$

Putting $u^i = \varphi_x^i$ we get

$$\frac{\partial b^j}{\partial \varphi^i} = \alpha_j(h_i) b^j,$$

whose solution is

$$b^j = B^j e^{\alpha_j(h_1) \varphi^1}, \quad (7)$$

The B^j being constants. (We sum over i in the exponential.)

We now turn to the integration of the second equation in (6). We assume the Lie algebra \mathfrak{g} to be semi-simple and that the vectors h_1 and σ_1 are chosen to be in the Weyl-Chevalley normal form (see [7], pp. 46-52). Let α_1 run over all roots in the vector figure, with $\alpha_{-1} = -\alpha_1^1$. We expand

$\sigma = c^k \sigma_k$. Then the second equation in (6) is

$$\varphi_{xt}^1 h_1 + b^j c^k [\sigma_k, \sigma_j] = 0.$$

Consequently

$$b^j c^k = 0 \text{ if } \alpha_j + \alpha_k \text{ is a non-zero root,}$$

and

$$\varphi_{xt}^1 = (b^1 c^{-1} - b^{-1} c^1) \quad (8)$$

In the case $sl(3, \mathbb{C})$ take $M = \{\alpha_1, \alpha_3, \alpha_{-2}\}$. We set $c^1 = c^3 = c^{-2} = 0$ and

choose b^1, b^3 , and b^{-2} to be non-zero. This leads to the equations

$$\begin{aligned} \varphi_{xt}^1 &= c^{-1} b^1 = c^{-1} e^{2(2\varphi_1 + \varphi_2 - \varphi_3)} \\ \varphi_{xt}^{-2} &= c^2 b^{-2} = c^2 e^{-2(\varphi_1 + 2\varphi_2 + \varphi_3)} \\ \varphi_{xt}^3 &= c^{-3} b^3 = c^{-3} e^{2(-\varphi_1 + \varphi_2 + 2\varphi_3)} \end{aligned} \quad (9)$$

(Recall that $h_{-1} = -h_1$ and $\varphi^{-1} = -\varphi^1$ so, for example,

$$\alpha_j(h_1)\varphi^1 = 2(\alpha_j(h_1)\varphi^1 + \alpha_j(h_2)\varphi^2 + \alpha_j(h_3)\varphi^3).$$

The associated isospectral problem is

$$D_x = \partial_x - 2 \sum_{i=1}^3 \varphi_x^i h_i - \zeta \begin{pmatrix} 0 & -1 & 0 & c^2 \\ c^{-1} & 0 & 0 & 0 \\ 0 & c^{-3} & 0 & 0 \end{pmatrix}$$

Equations (9) may be rewritten, as follows. Put

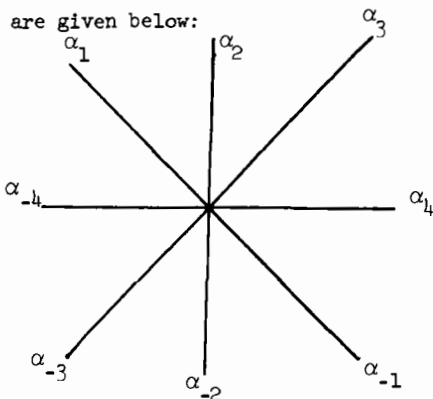
$$u = \alpha_1(h_1)\varphi^1 \quad v = \alpha_3(h_1)\varphi^1.$$

Then $u + v = (\alpha_1(h_1) + \alpha_3(h_1))\varphi^1 = \alpha_2(h_1)\varphi^1$, and equations (9) can be reduced to

$$\begin{aligned} u_{xt} &= 4c^{-1} e^u - 2c^2 e^{-(u+v)} - 2c^{-3} e^v \\ v_{xt} &= -2c^{-1} e^u - 2c^2 e^{-(u+v)} + 4c^{-3} e^v \end{aligned} \quad (10)$$

Now consider the algebra B_2 . Its vector figure and table of Cartan integers

are given below:



	h_1	h_2	h_3	h_4
α_1	2	2	0	-2
α_2	1	2	1	0
α_3	0	2	2	2
α_4	-1	0	1	2

$$a_i(h_j) = 2 \frac{\langle \alpha_i, \alpha_j \rangle}{\langle \alpha_j, \alpha_j \rangle}$$

We take $M = \{\alpha_1, \alpha_2, \alpha_3\}$, then we must set $c_j = 0$ whenever $\alpha_1 + \alpha_j \in P$ for some $\alpha_i \in M$. Consequently, we must have $c_{-2} = c_4 = c_{-4} = c_{-3} = c_{-1} = 0$. Our equations then are

$$\begin{aligned}\varphi_{xt}^1 &= -c_1 b^{-1} = -c_1 e^{-2(2\varphi_1 + 2\varphi_2)} \\ \varphi_{xt}^2 &= -c_2 b^{-2} = -c_2 e^{-2(\varphi_1 + 2\varphi_2 + \varphi_3)} \\ \varphi_{xt}^3 &= -c_3 b^{-3} = -c_3 e^{-2(2\varphi_2 + 2\varphi_3)}\end{aligned}$$

Again, defining

$$u = \alpha_1(h_j)\varphi^j = 2(2\varphi_1 + 2\varphi_2)$$

$$v = \alpha_3(h_j)\varphi^j = 2(2\varphi_2 + 2\varphi_3)$$

$$\frac{u+v}{2} = \alpha_2(h_j)\varphi^j = 2(\varphi_1 + 2\varphi_2 + \varphi_3)$$

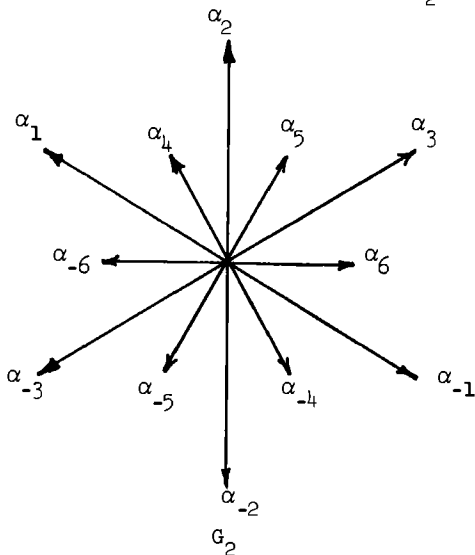
we have $\varphi_{xt}^1 = -c_1 e^{-u}$, $\varphi_{xt}^2 = -c_2 e^{-\frac{(u+v)}{2}}$, $\varphi_{xt}^3 = -c_3 e^{-v}$

and we arrive at the equations

$$\begin{aligned}u_{xt} &= c_1 e^{-u} + c_2 e^{-\frac{(u+v)}{2}} \\ v_{xt} &= c_2 e^{-\frac{(u+v)}{2}} + c_3 e^{-v}\end{aligned}$$

where c_1, c_2, c_3 are arbitrary constants.

Now we consider G_2 , the exceptional Lie algebra of rank 2. The vector figure and table of Cartan integers for G_2 is



	h_1	h_2	h_3	h_4	h_5	h_6
α_1	2	1	-1	3	0	-3
α_2	1	2	1	3	3	0
α_3	-1	1	2	0	3	3
α_4	1	1	0	2	1	-1
α_5	0	1	1	1	2	1
α_6	-1	0	1	-1	1	2

We take $M = \{\alpha_1, \alpha_2, \alpha_4, \alpha_5\}$ and set $c_{-1} = c_{-2} = c_3 = c_{-3} = c_{-4} = c_{-5} = c_6 = c_{-6} = 0$.

Our equations then are

$$\begin{aligned}\varphi_{xt}^1 &= -c_1^1 b^{-1} = -c_1^1 b^{-2(2\varphi_1 + \varphi_2 + 3\varphi_4)} \\ \varphi_{xt}^2 &= -c_2^2 b^{-2} = -c_2^2 e^{-2(\varphi_1 + 2\varphi_2 + 3\varphi_4 + 3\varphi_5)} \\ \varphi_{xt}^4 &= -c_4^4 b^{-4} = -c_4^4 e^{-2(\varphi_1 + \varphi_2 + 2\varphi_4 + \varphi_5)} \\ \varphi_{xt}^5 &= -c_5^5 b^{-5} = -c_5^5 e^{-2(\varphi_2 + \varphi_4 + 2\varphi_5)}\end{aligned}$$

as usual, we reduce these by introducing

$$\begin{aligned}u &= \alpha_1(h_1)\varphi^1 = 2(2\varphi_1 + \varphi_2 + 3\varphi_4) \\ v &= \alpha_5(h_1)\varphi^1 = 2(\varphi_2 + \varphi_4 + 2\varphi_5)\end{aligned}$$

Then since $\alpha_1 + \alpha_5 = 2\alpha_4$ and $\alpha_4 + \alpha_5 = \alpha_2$ we get

$$\alpha_4(h_1)\varphi^1 = \left(\frac{u+v}{2}\right) \quad \alpha_2(h_1)\varphi^1 = \left(\frac{u+3v}{2}\right)$$

so our equations become

$$\begin{aligned}u_{xt} &= -4c_1^1 e^{-u} - 2c_2^2 e^{-\left(\frac{u+3v}{2}\right)} - 6c_4^4 e^{-\left(\frac{u+v}{2}\right)} \\ v_{xt} &= -2c_2^2 e^{-\left(\frac{u+3v}{2}\right)} - 2c_4^4 e^{-\left(\frac{u+v}{2}\right)} - 4c_5^5 e^{-v}\end{aligned}$$

References

1. Ablowitz, M.J., Kaup, D.J., Newell, A.C., Segur, H., "Method for solving the sine-Gordon equation," *Phys. Rev. Lett.* 30, 1262-1264 (1973).
2. _____, "The inverse scattering transform-Fourier analysis for nonlinear problems," *Studies Appl. Math.* 53 (1974) 249-313.
3. Fordy, A.P., Gibbons, J., "A Class of Integrable Nonlinear Klein-Gordon Equations in Many Dependent Variables," *Commun. Math. Phys.* 77 (1980), 21-30.
4. Gardner, C.S., Greene, J.M., Kruskal, M.D., and Miura, R.M., "Korteweg - de Vries Equation and Generalizations, VI. Methods for Exact Solution." *Communications Pure and Applied Mathematics.* 27, 97-133.
5. Lax, P.D., "Integrals of Non-linear Equations of Evolution and Solitary Waves," *Comm. Pure Appl. Math.* 21 (1968), 467-490.
6. Mikhailov, A.V. Olshametsky, M.A., Peregomov, A.V. "Two Dimensional Generalized Toda Lattice," Preprint, Landau Institute for Theoretical Physics, Moscow, 1980.
7. Samuelson, H., Notes on Lie Algebras, Van Nostrand - Reinhold Mathematical Studies #23, New York, 1969.
8. Sattinger, D.H. "Prolongation Algebras and Gauge Theories for Isospectral Problems," Proceedings, 2nd. International Symposium on Dynamical Systems, Gainesville, Fla, 1981.
9. Wahlquist, H.D., and Estabrook, F.B., "Prolongation structures of nonlinear evolution equations," *Jour. Math. Phys.* 16 (1975), 1-7.
10. Zakharov, V.E. and Shabat, A.B., "Exact Theory of two dimensional self-focusing and one-dimensional self modulation of waves in nonlinear media," *Soviet Phys. JETP* 34 (1972), 62-69.

THE INVERSE MONODROMY TRANSFORM IS A CANONICAL TRANSFORMATION

H. Flaschka[†] and A.C. Newell^{*}

^{*}Clarkson College of Technology
Potsdam, NY 13676
U.S.A.

[†]University of Arizona
Tucson, AZ 85721
U.S.A.

I. INTRODUCTION AND GENERAL DISCUSSION

In earlier papers [1], [2], we used deformation theory to study the Painlevé equations which govern the self-similar solutions of the modified Korteweg deVries (MKdV) and sine-Gordon (SG) equations. In doing so we introduced the Inverse Monodromy Transform (IMT) which parallels the Inverse Scattering Transform (IST); whereas the latter is used to linearize the initial value problem for certain classes of non-linear evolution equations, the former allows us to find, by linear methods, an important class of solutions, the multiphase similarity solutions to these equations. They are to the standard (one-phase) similarity solutions what multisoliton or finite gap periodic solutions are to solitons and cnoidal waves.

We begin in section 2 by introducing the family of integrable evolution equations

$$P_{t_j} = P_j(P, P_x, \dots, P_{jx}), \quad (1.1)$$

(P_{jx} means the j^{th} derivative of P with respect to x) which are integrability conditions for a certain eigenvalue problem

$$V_x = (\zeta R + P(x, t_j))V, \quad -\infty < x < \infty, \quad (1.2)$$

and associated family

$$V_{t_j} = Q^{(j)}(\zeta; P, \dots, P_{(j-1)x})V. \quad (1.3)$$

Applying a further constraint

$$V_\zeta = S(\zeta; P, \dots, P_{(n-2)x})V \quad (1.4)$$

with S rational in ζ , defines a finite dimensional manifold of solutions common to a subset $j=0, 1, \dots, n-1$ of the equation set (1.1). The manifold is defined by

a nonautonomous, nonlinear ordinary differential equation in x , the coefficients depending on x and t_j , which is the analogue of the Lax-Novikov (LN) [3,4] equation that defines the finite-gap and multisoliton solutions of the members of (1.1). There are many parallels between the LN equation and its nonautonomous (NLN) counterpart.

In this paper, we

- (1) Introduce the NLN equation and the hierarchy of time flows with which it is associated.
- (2) Show that, with respect to a given symplectic form and Poisson bracket, the NLN equation (the x -flow) and the companion time flows are generated by a sequence of commuting Hamiltonians, closely related to those which generate the multiperiodic flows;
- (3) Define and illustrate how the IMT provides a mapping from the old coordinates P, P_x, \dots to new coordinates, the monodromy data M associated with the solution of (1.4), which admit trivial integration. In the case considered, S is polynomial in ζ of degree $(n-1)$ and so M consists of the Stokes multipliers $\{s_j\}_{j=1}^{2n}$ defined in connection with the rank n irregular singular point at $\zeta=\infty$.
- (4) Prove that the map is canonical and obtain expressions for the Poisson brackets of the Stokes multipliers.

Along the way, we

- (5) Develop expressions for the infinitesimal changes in the new coordinates in terms of a natural inner product between the old coordinates and a set of $(2n-2)$ vectors V_k which act as a basis in the space of dependent variables and which are formed from the "squared eigenfunctions". These expressions may be used to calculate the effects of perturbations on the NLN equations (for example, add a perturbation to the Painlevé equation; does it still retain its special properties? (see [1])) and provides a starting point for an attempt to prove a KAM (Kolmogoroff, Arnold, Moser) theorem - the flow does not, in general, take place on a compact manifold.
- (6) Develop expressions for the new coordinates themselves as inner products between the old coordinates and V_k .
- (7) Write expansions for the infinitesimal changes in the old coordinates and for the old coordinates themselves in terms of the basis vectors V_k . The latter provide expressions for solutions of the Painlevé and related equations in terms of contour integral representations.
- (8) The relation with analogous expressions which arise in connection with the inverse scattering transform are explored.

A more general discussion of the ideas presented here, together with additional material on multiperiodic systems, particle systems and perturbed systems will be given in a series of forthcoming papers [5]. Our work on these questions was stimulated by several works by Sato, Miwa, Jimbo, Mori and Ueno [6] relating to the correlation functions of exactly solvable models in statistical mechanics and quantum field theory. In [6], they discuss the Hamiltonian structure of deformation equations.

2. DEFINING THE EQUATIONS

The inverse scattering transform focuses principal attention on (1.2) where ζ is the eigenvalue, R a constant matrix and $P(x, t)$ a matrix of potentials which evolve in time $t = (t_0, t_1, \dots, t_{n-1})$ according to (1.1), $j=0, \dots, n-1$. From (1.2) and (1.3), we have

$$P_{t_j} - Q_x^{(j)} + [\zeta R + P, Q^{(j)}] = 0, \quad j = 0, 1, 2, \dots \quad (2.1a)$$

$$Q_{t_k}^{(j)} - Q_{t_j}^{(k)} + [Q^{(j)}, Q^{(k)}] = 0, \quad j, k = 0, 1, 2, \dots \quad (2.1b)$$

where $[\cdot, \cdot]$ is the commutator. We seek solutions of (2.1a) for $Q^{(j)}$ polynomial in ζ of degree j

$$Q^{(j)} = R \zeta^j + \sum_{k=0}^{j-1} Q_{j+1-k}^{(j)} \zeta^k, \quad (\text{the } Q_{j+1-k}^{(j)} \text{ are independent of } j) \quad (2.2)$$

and the $Q_{j+1-k}^{(j)}$ are solved from the coefficients of ζ^k (the part of $Q_{j+1-k}^{(j)}$ which does not commute with R) and of ζ^{k-1} (the part that does) in (2.1a) (see [7]). The last equation in the sequence gives the evolution equation (1.1). The part of $Q_{j+1-k}^{(j)}$ which commutes with R is determined up to a constant in x (it can depend on t_0, t_1, \dots, t_{j-1}) and we take this constant to be zero without loss of generality. If nonzero, we can make it zero by a transformation in the time coordinates

Example: Define

$$Y_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad (2.3)$$

$$R = -iY_1, \quad P = q(Y_2 + Y_3), \quad Q^{(3)} = (-i\zeta^3 - \frac{1}{2}q^2\zeta)Y_1 \\ + (\zeta^2 q - \frac{1}{4}q_{xx} + \frac{1}{2}q^3)(Y_2 + Y_3) + \frac{iq_x}{2}\zeta(Y_2 - Y_3),$$

and (1.1) is

$$q_{t_3} = -\frac{1}{4}(q_{xxx} - 6q^2q_x). \quad (2.4)$$

Finite-gap solutions of (1.1) are obtained by subjecting the vector V in (1.2) and (1.3) to a further (algebraic) constraint

$$\sum_{j=1}^{n-1} u_j V_{t_j} + u_0 V_x = \lambda V \quad (2.5a)$$

or

$$(Q - \lambda I)V = 0, \quad Q = \sum_{j=1}^{n-1} u_j Q^{(j)} + u_0(\zeta R + P). \quad (2.5b)$$

To motivate this choice, consider solutions of (2.6)

$$q = q(x - \zeta t_3), \quad X = x - \zeta t_3, \quad (2.6)$$

whence $q(X)$ satisfies

$$-\frac{1}{4}q_{xxx} + 3/2q^2q_x = cq_x. \quad (2.7)$$

In this case $Q^{(3)}$ is a function of ζ and X . Introduce the change of variables $X = x - \zeta t_3$, $T = t_3$ and (1.2), (1.3) with $j=3$ become

$$V_X = (\zeta R + P(X))V, \quad V_T = (Q^{(3)} + \zeta(\zeta R + P))V \quad (2.8)$$

respectively. Now set $V \rightarrow e^{\lambda T} V$ and obtain (2.5) with $u_3=1$, $u_0=c$, $u_j=0$ $j \neq 3$. Indeed the integrability condition of (1.2) and (2.5b) for the values chosen in the example is (2.7). In general, the integrability condition is

$$\sum_{j=1}^n u_j P_{t_j} + u_0 P_x = 0. \quad (2.9)$$

From (1.1), we see (2.9) is a nonlinear ordinary differential equation in x for the matrix P . The finite dimensional solution manifold defined by (2.12) is left invariant by the flows (1.1). By this we mean that finite-gap solutions considered as function of x satisfy (2.9) for all values the time parameters. It is well known (and details have been worked out in several cases [8]) that as functions of x , it is a completely integrable autonomous Hamiltonian system.

Solutions of (2.9) (as well as their time evolution under (1.1)) can be constructed explicitly; they are abelian functions. They are closely connected with the Riemann surface R defined by the polynomial relation $\Gamma(\zeta, \lambda) = 0$ between ζ and λ which comes about by the vanishing of the determinant of $Q - \lambda I$. The points of the Riemann surface parametrize $V(x, t)$ in (1.2), (1.3) and (2.5b). By proper normalization of V as eigenvector of (2.5b), one can arrange that it satisfies (1.2) and (1.3) and at the point(s) at infinity on R ,

$$V \sim [1 + O(\frac{1}{k})] \exp(kx + \sum_{j=1}^n \Omega_j(k) t_j) \quad (2.10)$$

k being a local parameter. The $\Omega_j(k)$ are polynomials, usually the dispersion relations of the linearized equations (1.1). In addition, V as function of (ζ, λ) on R has poles μ_s which are independent of x and t_j . Specification of the surface R , the poles μ_s , the asymptotics (2.10) and of a certain normalization condition determines V uniquely. Knowing V one can, by differentiation with respect to x , recover P from (1.2).

Now let us turn to a new class of solutions defined by adding another kind of constraint on the function $V(x, t, \zeta)$. This time we will ask it to satisfy an ordinary differential equation in ζ ,

$$\zeta V_\zeta = \sum_j t_j V_{t_j} + x V_x, \quad (2.11a)$$

$$= (\sum_{j=1}^n j t_j Q^{(j)} + x(\zeta R + P)) V. \quad (2.11b)$$

Equation (2.11) has coefficients which are rational functions of ζ ; all dependence on x, t_j is only parametric as far as ζ is concerned. The integrability condition on (2.11) and (1.2) gives us that (we omit the t_0 flow)

$$\sum_{j=1}^n j t_j P_{t_j} + (xP)_x = 0 \quad (2.12)$$

which shows that P is a function of the phases $\frac{x}{(nt_n)^{1/n}}, \frac{t_j}{(nt_n)^{1/n}}, j=1, \dots, n-1$.

Equation (2.12), the analogue of (2.9), is an ordinary differential equation in x of order n with coefficients which depend on t_j . If we begin at time $\tilde{t}^{(0)}$ with a solution of (2.12) and then let it evolve for a time $\tilde{t} - \tilde{t}^{(0)}$ in the flows (1.1), then at time \tilde{t} , it will again satisfy (2.12) with the coefficients t_1, \dots, t_n evaluated at the new time

The choice of (2.11) can be motivated with the aid of the example (2.4) which has the self similar solution

$$q = \frac{1}{(3t_3)^{1/3}} f\left(\frac{x}{(3t_3)^{1/3}}\right) \quad (2.13a)$$

and $f(x)$, $x = x/(3t_3)^{1/3}$, satisfies the Painlevé equation of the second kind

$$\frac{1}{4}(f_{xx} - 2f^3) = xf - v. \quad (2.13b)$$

If we choose q to have the form (2.13a) in (1.2) and (1.3) with $j=3$, then V is a function of x, t_3 and ζ only through the combinations $x = x/(3t_3)^{1/3}$, $\zeta = \zeta(3t_3)^{1/3}$. Then (1.3), with $j=3$, becomes (2.11b) with $t_j=0$ $j \neq 3$. In the context of the example, the integrability condition of (2.11b) and (1.2) is (2.13b).

Another class of solutions to (1.1) can be found by adding the constraint

$$v_\zeta = \left(\sum_{j=1}^n \beta_j Q^{(j-1)+xR} \right) V \quad (2.14)$$

which leads to the integrability condition

$$\sum_{j=1}^n \beta_j P_{t_{j-1}} + x P_{t_0} = 0. \quad (2.15)$$

Recall $P_{t_0} = [R, P]$ is the scaling flow and $P_{t_1} = P_x$ is translation. In order that (2.14) and (1.3) are compatible (examine the asymptotic behavior of V as $\zeta \rightarrow \infty$), $\beta_j = jt_j$. Equation (2.15) then implies that $P(t_{n-1}, t_{n-2}, \dots, t_2, x, t_0)$ is a function only of the n phases

$$P = P(\tau_{n-2}, \dots, \tau_1, \tau_0) \quad (2.10)$$

where $\{\tau_j\}_0^{n-2}$ are the integration constants defined by solving the equations

$$\frac{dt_{n-1}}{nt} = \frac{dt_{n-2}}{(n-1)t_{n-1}} = \dots = \frac{dx}{2t_2} = \frac{dt_0}{x} \quad (2.17)$$

where without loss of generality we can take $nt_0=1$. It is this class of solutions which we discuss when (1.2) is the Zakharov-Shabat system $\zeta R + P = \begin{pmatrix} -1/\zeta & q \\ r & i\zeta \end{pmatrix}$.

By taking the coefficient matrix in the constraint equation to be a more complicated combination of rational functions of ζ , we can of course build more elaborate classes of solutions. In [2], we briefly indicated how to include a combination of solitons and self-similar solutions. In [5], we describe how to include the x dependence of the NLN system in different ways.

3. THE ZAKHAROV-SHABAT SYSTEM

3a. The equations: Here (1.2), (1.3) and the constraint equation (2.14) are

$$v_x = \begin{pmatrix} -i\zeta & q \\ r & i\zeta \end{pmatrix} v, \quad (3.1a)$$

$$v_{t_j} = \begin{pmatrix} -a(j) & b(j) \\ c(j) & a(j) \end{pmatrix} v, \quad (3.1b)$$

$$\zeta V_{\zeta} = \begin{pmatrix} -a & b \\ c & a \end{pmatrix} V, \quad (3.1c)$$

the integrability conditions for which are

$$a_x = rb - qc + i\zeta, \quad b_x + 2i\zeta b = 2qa, \quad c_x - 2i\zeta c = -2ra, \quad (3.2)$$

$$a_x^{(j)} = rb^{(j)} - qc^{(j)}, \quad b_x^{(j)} + 2i\zeta b^{(j)} = q_{t_j} + 2qa^{(j)}, \quad c_x^{(j)} - 2i\zeta c^{(j)} = r_{t_j} - 2ra^{(j)}, \quad (3.3)$$

$$a_{t_j} = c^{(j)}b - b^{(j)}c + \zeta a_{\zeta}^{(j)}, \quad b_{t_j} + 2a^{(j)}b = 2b^{(j)}a + \zeta b_{\zeta}^{(j)}, \quad c_{t_j} - 2a^{(j)}c = -2c^{(j)}a + \zeta c_{\zeta}^{(j)}. \quad (3.4)$$

3b. The Solutions To (3.2): We seek solutions to (3.2) polynomial in ζ . Set

$$a = \sum_{k=-\infty}^n a_{n+1-k} \zeta^k, \quad b = \sum_{k=-\infty}^{n-1} b_{n+1-k} \zeta^k, \quad c = \sum_{k=-\infty}^{n-1} c_{n+1-k} \zeta^k. \quad (3.5)$$

We find $a_1 = \alpha_1$ constant, a_2 can be omitted. The ordinary differential equations (2.12) or (2.15) are formed by terminating the series at ζ^0 and ζ^1 respectively. We follow the latter course and then the equations are

$$b_{n+1} = c_{n+1} = 0 \quad (3.6)$$

where $b_j, c_j, j=2, \dots, n+1$ are defined

$$b_{j+1} = \frac{1}{2}b_{jx} - ia_j q + xq\delta_{j,n}, \quad c_{j+1} = -\frac{1}{2}c_{jx} - ia_j r + r\delta_{j,n}, \quad j=1, \dots, n. \quad (3.7)$$

The coefficients $a_j, j=3, \dots, n$, are found by noting that the function

$$\Omega = a^2 + bc - 2i\zeta r^x \text{ad}x \quad (3.8)$$

is independent of x . We set this equal to

$$\Omega = \left(\alpha_1 \zeta^n + \alpha_3 \zeta^{n-2} + \dots + \alpha_n \zeta + \alpha_{n+1} + \frac{\alpha_{n+2}}{\zeta} + \dots \right)^2, \quad (3.9)$$

and equate the powers of $\zeta^{n+\ell}$ in (3.8). For $\ell=n-2$ to $\ell=1$, this equation defines $a_{n+1-\ell}$ in terms of the b_j, c_j already determined to this level from (3.7) and $\alpha_{n+1-\ell}$ is effectively the additive constant associated with integrating a_x in (3.2). For $b=0$ to $b=-n+2$, this equation determines $\alpha_{n+1-\ell}$ ($\alpha_{n+1}, \dots, \alpha_{2n-1}$) in terms of the already known $a_j, b_j, c_j, j \leq n$ and $(x, \alpha_3, \alpha_4, \dots, \alpha_n)$. For $\ell < -n+2$, we simply choose $\alpha_{n+1-\ell}$ such that the coefficients of $\zeta^{n+\ell}, \ell < -n+2$ are zero. These choices play no role in the later analysis. We now list the first six coefficients.

	a	b	c
1.	1	0	0
2.	0	$-i\alpha_1 q$	$-i\alpha_1$
3.	$\alpha_3 + \alpha_1 q r / 2$	$\frac{\alpha_1}{2} q_x$	$-\frac{\alpha_1}{2} r_x$
4.	$\alpha_4 + \frac{i\alpha_1}{4} (r q_x - r_x q)$	$\frac{i\alpha_1}{4} (q_{xx} - 2q^2 r) - i\alpha_3 q$	$\frac{i\alpha_1}{4} (r_{xx} - 2qr^2) - i\alpha_3 r$
5.	$\alpha_5 + \alpha_3 \frac{qr}{2}$	$-\frac{\alpha_1}{8} (q_{xxx} - 6qrq_x) + \frac{\alpha_3}{2} q_x$	$\frac{\alpha_1}{8} (r_{xxx} - 6qrr_x) - \frac{\alpha_3}{2} r_x$
	$-\frac{\alpha_1}{8} (r q_{xx} + q r_{xx} - r_x q_x - 3q^2 r^2)$	$-i\alpha_4 q$	$-i\alpha_4 r$
6.	$\alpha_6 + \alpha_4 \frac{qr}{2} + \frac{i\alpha_3}{4} (r q_x - r_x q)$	$-\frac{i}{16} \alpha_1 (q_{xxx} - 6qrq_x)_x$	$-\frac{i}{16} \alpha_1 (r_{xxx} - 6qrr_x)_x$
	$-\frac{i}{16} \alpha_1 (r q_{xxx} - q r_{xxx} - r_x q_{xx})$	$+\frac{i\alpha_1}{8} (r q_{xx} + q r_{xx} - q_x r_x - 3q^2 r^2) q$	$+\frac{i\alpha_1}{8} (r q_{xx} + q r_{xx} - q_x r_x - 3q^2 r^2) r$
	$+q_x r_{xx} - 6qr(r q_x - r_x q)$	$+\frac{i\alpha_3}{4} (q_{xx} - 2q^2 r) + \frac{\alpha_4}{2} q_x - i\alpha_3 q$	$+\frac{i\alpha_3}{4} (r_{xx} - 2qr^2) - \frac{\alpha_4}{2} r_x - i\alpha_3 r$

(3.10)

3c. The Solutions of (3.3): We seek polynomial solutions to (3.3) in the form

$$a(j) = i\zeta^j + \sum_{k=0}^{j-1} a_{j+1-k} \zeta^k, \quad b(j) = \sum_{k=0}^{j-1} b_{j+1-k} \zeta^k, \quad c(j) = \sum_{k=0}^{j-1} c_{j+1-k} \zeta^k. \quad (3.11)$$

It is simple to verify that $a_{n+1-k}^{(j)}$, $b_{n+1-k}^{(j)}$, $c_{n+1-k}^{(j)}$ are simply a_{n+1-k} , b_{n+1-k} , c_{n+1-k} when $n=j$, $\alpha_1=i$, $\alpha_k=0$, $k=3--j$. For example, $a^{(3)} = i\zeta^3 + \frac{iqr}{2}\zeta - \frac{1}{4}(r q_x - r_x q)$. The ζ^0 balance in (3.3) gives the evolution equations for q and r as functions of t_j . Because of the normalization of $a^{(j)}$, the compatibility of (3.2 and (3.3) requires that

$$\alpha_{n+1-j} = i j t_j, \quad j = 2, --n. \quad (3.12)$$

We take $t_{n-1}=0$ and omit the t_{n-1} flow. We will show very shortly how this dependence can be reincorporated. In essence, the t_{n-1} flow is simply a restatement of the NLN equations (3.6) once the particular choice of scaling in $t_0, x, t_2, \dots, t_{n-1}$ is made. We can take $nt_n=1$. Also we will take $t_1=0$ as this flow simply mimics the x -flow. We will often for convenience call x by t_j . We are left with the x -flow given by (3.6) and the time flows q_{t_j} , r_{t_j} , $j=0, 2, --n-2$. We list and discuss the cases $n=3, 4$.

$n=3$: The x -flow is

$$b_{3x} = 2(a_3 + ix)q, \quad c_{3x} = -2(a_3 + ix)r$$

or

$$\frac{i}{2}(q_{xx} - 2q^2 r) = 2ixq, \quad \frac{i}{2}(r_{xx} - 2qr^2) = 2ixr. \quad (3.13)$$

The only time flow is

$$q_{t_0} = -2iq, r_{t_0} = 2ir. \quad (3.14)$$

Now let us comment on what this solution has to do with solutions of

$$q_{t_2} = \frac{1}{2}(q_{xx} - 2q^2r), r_{t_2} = -\frac{1}{2}(r_{xx} - 2qr^2). \quad (3.15)$$

If we had left $\alpha_2 \neq 0$ and indeed chosen it to be $2it_2$, then (3.13) would read with $\alpha_1 = 1$,

$$q_{t_2} + 2t_2 q_{t_1} + xq_{t_0} = 0, r_{t_2} + 2t_2 r_{t_1} + xr_{t_0} = 0 \quad (3.16)$$

which means that q and r are functions of

$$\tau_1 = x - t_2^2, \tau_0 = t_0 + \frac{2}{3}t_2^3 - xt_2 \quad (3.17)$$

Now impose (3.17), and since (3.14) holds, we have

$$\begin{aligned} q(t_0, x, t_2) &= e^{-2i(t_0 + \frac{2}{3}t_2^3 - xt_2)} f(x - t_2^2) \\ r(t_0, x, t_2) &= e^{2i(t_0 + \frac{2}{3}t_2^3 - xt_2)} g(x - t_2^2) \end{aligned} \quad (3.18)$$

which when substituted into (2.15) has f and g satisfying (3.13) with $\alpha_1 = i$ and x replaced by $x - t_2^2$.

Thus in what follows we simply ignore the equation for $q_{t_{n-1}}, r_{t_{n-1}}$ and find q, r as functions of $t_0, x, t_2, \dots, t_{n-2}$. Then to incorporate t_{n-1} simply replace $t_0, x, t_2, \dots, t_{n-2}$ by $\tau_0, \tau_1, \dots, \tau_{n-2}$ which are found by integrating

$$dt_{n-1} = \frac{dt_{n-2}}{(n-1)t_{n-1}} = \frac{dx}{2t_2} = \frac{d\tau_0}{x}. \quad (3.19)$$

n=4 The x -flows are

$$b_{4x} = 2(a_4 + ix)q, c_{4x} = -2(a_4 + ix)r$$

or

$$\begin{aligned} -\frac{1}{4}(q_{xxx} - 6qrq_x) - i\alpha_3 q_x - 2ixq &= 0 \\ -\frac{1}{4}(r_{xxx} - 6qrq_x) - i\alpha_3 r_x + 2ixr &= 0. \end{aligned} \quad (3.20)$$

The associated time flows are

$$q_{t_0} = -2iq, r_{t_0} = 2ir \quad (3.21)$$

$$q_{t_2} = \frac{1}{2}(q_{xx} - 2q^2r), r_{t_2} = -\frac{1}{2}(r_{xx} - 2qr^2). \quad (3.22)$$

Again we may also include solutions to

$$q_{t_3} = -\frac{1}{4}(q_{xxx} - 6qrq_x), \quad r_{t_3} = -\frac{1}{4}(r_{xxx} - 6qrr_x) \quad (3.23)$$

by replacing the solutions

$$q(t_0, x, t_2), \quad r(t_0, x, t_2)$$

by

$$q(t_0 - \frac{3t_3^4}{4} + \frac{t_2 t_3^2}{2} - xt_3, \quad x + t_3^3 - t_2 t_3, \quad t_2 - \frac{3t_3^2}{2})$$

and

$$r(t_0 - \frac{3t_3^4}{4} + \frac{t_2 t_3^2}{2} - xt_3, \quad x + t_3^3 - t_2 t_3, \quad t_2 - \frac{3t_3^2}{2}).$$

In the corresponding periodic problem, ζV_r in (3.1c) is replaced by λV . This removes the $i\zeta x$ term in a , the function Ω is $a^2 + bc$ and the identification (3.12) is no longer necessary; the α_j are constants. The Lax-Novikov equations are

$$b_{nx} = 2a_n q, \quad c_{nx} = -2a_n r. \quad (3.24)$$

3d. The Hamiltonian Structure: The function Ω generates the Hamiltonian for each of the flows. In particular

$$H_j = 4\alpha_{n+1+j}, \quad j = 0, \dots, n-2. \quad (3.25)$$

The conjugate variables are

$$\bar{b}_j, \quad \bar{c}_{n+2-j} \quad j=2, \dots, n \quad (3.26)$$

where \bar{b}_j, \bar{c}_j are the coefficients of ζ^{-j+2} , $j=2, \dots, n$ in the asymptotic expansions for

$$\zeta^{n-1} \sqrt{\frac{b}{\alpha_1 \zeta^n}}, \quad \zeta^{n-1} \sqrt{\frac{c}{\alpha_1 \zeta^n}}, \quad \alpha = \alpha_1 \zeta^n + \alpha_3 \zeta^{n-2} + \dots \quad (3.27)$$

as $\zeta \rightarrow \infty$. We can identify the correct choice of conjugate variables from three sources:

- (i) they are the same as those of the periodic problem (the detailed analysis leading to these choices is given in [6];
- (ii) the Hamiltonian H_0 which generates the scaling flow must decompose into a sum of products of conjugate variables;
- (iii) the choices are necessary in order to define certain inner products correctly.

For now, we will simply take (3.27) as given. The Hamiltonians contain the independent variables x, t_2, \dots, t_{n-2} . We make the system autonomous by including these as dependent variables and adding conjugates $T_1 = X, T_2, \dots, T_{n-2}$ which are defined from the term $\int^x a_j dx$ in $j=3, \dots, n$ in such a way that

$$H_j = T_j + \hat{H}_j(\bar{b}_k, \bar{c}_k, x, t_2, \dots, t_{n-2}). \quad (3.28)$$

From this choice we see that

$$t_j = \frac{\partial H_j}{\partial T_j} = 1, \quad t_k = \frac{\partial H_j}{\partial T_k} = 0, \quad k \neq j (t_1 = x),$$

and so t_j is the "time" variable for the flow generated by H_j .

Let us look in detail at some examples.

$$n=3: \quad \alpha_1 = i, \quad \alpha_3 = 0.$$

$$H_0 = -2i(b_2 c_3 + b_3 c_2)$$

$$H_1 = -2i(b_3 c_3 - \frac{b_2^2 c_2^2}{4} - x b_2 c_2 + \int^x b_2 c_2 dx).$$

The conjugate variables are

$$b_2, b_3, x \text{ and } c_3, c_2, X = -2i \int^x b_2 c_2 dx.$$

It is easy to verify that H_0, H_1 generate the flows (3.14) and (3.13), respectively. Also, both H_0 and H_1 are constants with respect to t_0 and x .

$$n=4: \quad \alpha_1 = i, \quad \alpha_3 = 2it_2, \quad \alpha_4 = 0.$$

$$H_0 = -2i(\bar{b}_2 \bar{c}_4 + \bar{b}_3 \bar{c}_3 + \bar{b}_4 \bar{c}_2)$$

$$\text{where } \bar{b}_2 = b_2, \quad \bar{b}_3 = b_3, \quad \bar{b}_4 = b_4 - \frac{a_3 + \alpha_3}{4i} b_2 = b_4 - \frac{b_2^2 c_2}{8} - t_2 b_2$$

and the c 's are defined analogously.

$$H_1 = -2i\{\bar{b}_3(\bar{c}_4 + \frac{\bar{b}_2 \bar{c}_2^2}{8} + t_2 \bar{c}_2) + \bar{c}_3(\bar{b}_4 + \frac{\bar{b}_2 \bar{c}_2^2}{8} + t_2 \bar{b}_2) \\ + \frac{1}{2}(2it_2 + \frac{1}{2}\bar{b}_2 \bar{c}_2)(\bar{b}_2 \bar{c}_3 + \bar{b}_3 \bar{c}_2) - x \bar{b}_2 \bar{c}_2 + \int^x \bar{b}_2 \bar{c}_2 dx\}$$

$$H_2 = -2i\{(\bar{b}_4 + \frac{\bar{b}_2 \bar{c}_2^2}{8} + t_2 \bar{b}_2)(\bar{c}_4 + \frac{\bar{b}_2 \bar{c}_2^2}{8} + t_2 \bar{c}_2) - x(\bar{b}_2 \bar{c}_3 + \bar{b}_3 \bar{c}_2) \\ + \int^x (\bar{b}_2 \bar{c}_3 + \bar{b}_3 \bar{c}_2) - 2t_2(\bar{b}_2 \bar{c}_4 + \bar{b}_3 \bar{c}_3 + \bar{b}_4 \bar{c}_2)\}.$$

Note that the last term in H_2 comes from $-4i\alpha_3\alpha_5$ appearing in the coefficient of c^2 in (3.8). The conjugate pairs are (\bar{b}_2, \bar{c}_4) , (\bar{b}_3, \bar{c}_3) , (\bar{b}_4, \bar{c}_2) , $(x, X = -2i \int^x \bar{b}_2 \bar{c}_2 dx)$,

$$(t_2, T_2 = -2i \int^x (\bar{b}_2 \bar{c}_3 + \bar{b}_3 \bar{c}_2) dx).$$

In particular the form of T_2 may be obtained by noting that, since H_1 generates the x flow,

$$\frac{dT_2}{dx} = -\frac{\partial H_1}{\partial t_2} = -2i(\bar{b}_2 \bar{c}_3 + \bar{b}_3 \bar{c}_2).$$

In this way, the form of T_k , $k=2, \dots, n-2$ can always be found from the corresponding H_1 .

Before we continue, a convention or notation. $\frac{df}{dt_j}$ means differentiate all variables in f which depend on t_j , including t_j itself, keeping $t_k \neq t_j$ ($t_1=x$) constant. $\frac{\partial f}{\partial t_j}$ means we differentiate f only with respect to explicit t_j dependence.

For the multiperiodic case, we define H_0, H_1, \dots, H_{n-2} by leaving out the contributions from $2i\zeta^j \lambda dx$. We also leave the α_j as constants. Then the conjugate variables are simply (3.26), the equations are the same with α replaced by constant α_n .

3e. Poisson Brackets And The Commutability of $\{H_j\}_{j=0}^{n-2}$.

It is natural to define the Poisson bracket ($t_1=x, T_1=X$)

$$\begin{aligned} [F, G] = & \sum_{n \neq 2}^n \left(\frac{\partial F}{\partial \bar{b}_j} \frac{\partial G}{\partial \bar{c}_{n+2-j}} - \frac{\partial F}{\partial \bar{c}_{n+2-j}} \frac{\partial G}{\partial \bar{b}_j} \right) \\ & + \sum_{j=0}^{n-2} \left(\frac{\partial F}{\partial t_j} \frac{\partial G}{\partial T_j} - \frac{\partial F}{\partial T_j} \frac{\partial G}{\partial t_j} \right). \end{aligned} \quad (3.29)$$

The first term is the Poisson bracket for the periodic problem and we write this $[F, G]_p$. Note that if

$$\nabla = \left(\frac{\partial}{\partial \bar{c}_2}, \frac{\partial}{\partial \bar{b}_2}, \frac{\partial}{\partial \bar{c}_3}, \frac{\partial}{\partial \bar{b}_3}, \dots, \frac{\partial}{\partial \bar{c}_n}, \frac{\partial}{\partial \bar{b}_n} \right)^T, \quad (3.30)$$

$$J = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & \dots & \dots & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & \dots & 0 \end{pmatrix} \quad (3.31)$$

then

$$[F, G]_p = \langle \nabla F, J \nabla G \rangle, \quad (3.32)$$

where $\langle u, v \rangle$ is $\sum_{j=2}^n u_j v_j$, the usual inner product. Also, if

$$F = F(\bar{b}_j, \bar{c}_j, t_j, T_j),$$

then

$$\frac{dF}{dt_k} = [F, H_k]. \quad (3.33)$$

In [5], we show that for each $j, k = 0, \dots, n-2$,

$$[H_j, H_k] = 0. \quad (3.34)$$

Thus each H_j is a constant of the motion for each of the flows generated by the

other H_j . We therefore have $(n-1)$ independent constants of the motion in involution. The periodic system, which has $(2n-2)$ dependent variables, is then completely integrable. The multiphase self-similar solutions on the other hand are $4n-4$ dimensional and thus we need another $(n-1)$ independent constants of the motion. These new constants are introduced in the next section.

4. The Inverse Monodromy Transform (IMT):

Consider (3.1c),

$$\zeta V_\zeta = \begin{pmatrix} -a & b \\ c & a \end{pmatrix} V \quad (4.1)$$

with a, b, c , given by (3.5) with $a_{n+1} = b_{n+1} = c_{n+1} = 0$. Then $\zeta = \infty$ is an irregular singular point of rank n and the fundamental solution matrix

$$\Phi(x, t_j; \zeta) = (\psi, \bar{\psi}) \quad (4.2)$$

admits the formal asymptotic expansion (see Appendix I) as $\zeta \rightarrow \infty$,

$$\Phi \sim \tilde{\Phi} = \left(I + \sum_{s=1}^{\infty} \frac{C_s}{\zeta^s} \right) e^{-\Phi} \begin{pmatrix} e^{-\theta} & 0 \\ 0 & e^{\theta} \end{pmatrix} \quad (4.3)$$

where

$$\theta = \frac{i\zeta^n}{n} + it_{n-2}\zeta^{n-2} + \dots + it_j\zeta^j + i\zeta x + it_0 + \frac{i\ell_0}{4} \ln \zeta + o(1),$$

$$\Phi = o(1), \quad H_0 = +4\alpha_{n+1} \quad (4.4)$$

$\tilde{\Phi}$ has been normalized so as to satisfy (3.1a) and (3.1b). Now, it is known that in general Φ will not have the asymptotic expansion $\tilde{\Phi}$ in every neighborhood of $\zeta = \infty$. This neighborhood naturally divides into $2n$ equal sectors separated by rays, called anti-Stokes lines, on which $\operatorname{Re} \theta = \operatorname{Re} i\zeta^n = 0$. Define S_k as $\{\zeta; |\zeta| > \rho, \frac{\pi}{n}(k-1) \leq \operatorname{Arg} \zeta < \frac{\pi}{n}k\}$

and R_k as $\operatorname{Arg} \zeta = \frac{\pi}{n}(k-1)$. Let $\Phi = (\psi, \bar{\psi})$ be a fundamental solution matrix of (4.1) with asymptotic behavior $\tilde{\Phi}$ in S_1 . The solution $\tilde{\psi} \begin{pmatrix} 0 \\ 1 \end{pmatrix} e$ is recessive (asymptotically decaying whereas $\psi \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-\theta}$ is dominant (it is unique as it is defined on the initial ray $\zeta=0$ where θ is imaginary). Continuing to sector S_2 , the recessive solution $\bar{\psi}$ becomes the dominant solution $\bar{\psi}_2$ in that sector; however, in order that the dominant solution ψ in S_1 become the recessive solution ψ_2 in S_2 , we must add a constant factor (the Stokes multiplier) times the recessive solution $\bar{\psi}$. In general, the fundamental solution matrix Φ_{j+1} which has asymptotic behavior $\tilde{\Phi}$ in S_{j+1} is related to its preceding neighbor by

$$\Phi_{j+1} = \Phi_j M_j, \quad (4.5)$$

where $M_j = \begin{pmatrix} 1 & s_j^1 \\ s_j & 1 \end{pmatrix}$. If ψ_j is dominant, then $s_j^1 = 0$, $s_j \neq 0$. If ψ_j is recessive and

$\bar{\psi}_j$ dominant, then the nonzero off-diagonal element of M_j occupies the (1,2) position.

We call the set of Stokes multipliers $\{s_j\}_{j=1}^{2n}$ the monodromy data M for (4.1). Since the total monodromy around $\zeta=\infty$ is the identity (as it equals that about $\zeta=0$, an ordinary point), only $(2n-2)$ of the set M are independent. For reasons of symmetry we will choose to work with the set $\{s_k\}$ for $k \in \mathbb{Z} = \{1, 2, \dots, -n-1, n+1, \dots, -2n-1\}$.

One of the central results of our previous papers is: given ϕ_j , $j=1, \dots, 2n$ with asymptotic behaviors Φ which satisfy (3.1a), (3.1b), (3.1c), and $M_j(s_j)$ defined by (4.5), then the Stokes multipliers are constants, independent of t_0, x, \dots, t_{n-2} . This result is again proved in Section 5.

We therefore have a map from old variables

$$q(\bar{b}_2, \dots, \bar{b}_n, x, t_2, \dots, t_{n-2}), p(\bar{c}_n, \dots, \bar{c}_2, X, T_2, \dots, T_{n-2})$$

to new variables

$$Q(f_1(s_k), \dots, f_{n-1}(s_k), x_1 t, \dots, t_{n-2}), P(g_1(s_k), \dots, g_{n-1}(s_k), H_1, \dots, H_{n-2})$$

where the functions f_j and g_j are functions of the Stokes multipliers $s_k, k \in \mathbb{Z}$. In the new variables the equations are

$$\begin{aligned} s_j, t_k &= 0, \quad j \in \mathbb{Z}, \quad k = 0, 1, \dots, -n-2, t_1 = x. \\ H_j, t_k &= 0, \quad j = 0, \dots, -n-2, \quad k = 0, 1, \dots, -n-2. \\ t_j, t_k &= \delta_{jk} \quad j, k = 0, \dots, -n-2. \end{aligned} \quad (4.6)$$

Our next goal is to show that this map is canonical. In order to do this, it is necessary to express the infinitesimal variations $\delta Q, \delta P$ in terms of $\delta q, \delta p$.

5. INFINITESIMAL VARIATIONS AND THE t_j DEPENDENCE OF s_k .

5a. Infinitesimal variations: Take the infinitesimal variation of

$$\zeta \Phi = \begin{pmatrix} -a & b \\ c & a \end{pmatrix} \Phi, \Phi = \begin{pmatrix} \psi_1 & \bar{\psi}_1 \\ \psi_2 & \bar{\psi}_2 \end{pmatrix} \quad (5.1)$$

and solve by variation of parameters. Integrate the result between ζ_1 , which lies near $\zeta=\infty$ on R_1 , and ζ_2 , defined similarly. Since Φ is entire, any path will do. We find

$$\Phi^{-1} \delta \Phi \Big|_{\zeta_2} - \Phi^{-1} \delta \Phi \Big|_{\zeta_1} = \int_{\zeta_1}^{\zeta_2} \frac{1}{\zeta} \Phi^{-1} \begin{pmatrix} -\delta a & \delta b \\ \delta c & \delta a \end{pmatrix} \Phi d\zeta. \quad (5.2)$$

Now use (4.3) and the facts that on $R_1, \Phi \sim \tilde{\Phi}$, on $R_2, \Phi \sim M_1^{-1}$, and compare the (2,1) elements of (5.2) to obtain

$$-\delta s_1 - 2i s_1 \left(\frac{n-2}{2} \delta t_j \zeta_2^j + \delta x \zeta_2 + \delta t_0 \right) - 2 s_1 \delta \alpha_{n+1} \ln \zeta_2 = \int_{\infty_1}^{\zeta_2} (2 \delta a \psi_1 \psi_2 + \delta b \psi_2^2 + \delta c \psi_1^2) \frac{d\zeta}{\zeta}, \quad (5.3)$$

where we have let $\zeta_1 \rightarrow \infty_1$, the point at ∞ on R_1 . We cannot set $\zeta_2 = \infty_2$, for neither side

of (5.3) converges. However, we make use of the following identities derived from (3.1):

$$\frac{1}{\zeta} (2a_x \psi_1 \psi_2 - b_x \psi_2^2 + c_x \psi_1^2) = \frac{d}{d\zeta} (2i\zeta \psi_1 \psi_2 - q \psi_2^2 + r \psi_1^2), \quad (5.4a)$$

$$\frac{1}{2} (2a_{t_j} \psi_1 \psi_2 - b_{t_j} \psi_2^2 + c_{t_j} \psi_1^2) = \frac{d}{d\zeta} (2a^{(j)} \psi_1 \psi_2 - b^{(j)} \psi_2^2 + c^{(j)} \psi_1^2).$$

Also from (A.2), we may show that the arguments of the right hand sides of (5.6a) and (5.6b) have asymptotic behaviors

$$-2i\zeta_2 s_1 + o(1) \text{ and } -2i\zeta_2^j s_1 + o(1) \text{ as } \zeta_2 \rightarrow \infty \text{ on } R_2$$

respectively. Therefore we may write

$$-\delta s_1 = \int_{\infty_1}^{\zeta_2} (2da \psi_1 \psi_2 - db \psi_2^2 + dc \psi_1^2) \frac{d\zeta}{\zeta} + 2s_1 \delta \alpha_{n+1} \ln \zeta_2 \quad (5.5)$$

where $da = \delta a - \sum_{j=0}^{n-2} a_{t_j} \delta t_j$, ($t_1 = x$) and db, dc are defined similarly.

The integral on the right hand side of (5.5) has an asymptotic expansion which consists of (a) terms like $e^{\pm 2\theta} \zeta^p$ when $p < n-1$, which are integrable along the rays R_1 and R_2 and (b) a term proportional to $1/\zeta$ on R_2 . When integrated, this term is exactly cancelled by $2s_1 \delta \alpha_{n+1} \ln \zeta_2$. Thus the limit $\zeta_2 \rightarrow \infty$ along R_2 may be taken and $-\delta s_1$ is well defined.

The calculation may be repeated for any sector. In the odd numbered sectors, one compares the (2,1) elements of (5.2); for the even ones compare the (1,2) elements. This results in a change of sign. If $\zeta_R(\infty_R)$ is a point which tends to $\zeta = \infty$ along R_k , then,

$$(-1)^k \delta s_k = \int_{\infty_k}^{\zeta_{k+1}} (2da \psi_1 \psi_2 - db \psi_2^2 + dc \psi_1^2) \frac{d\zeta}{\zeta} + 2s_k \delta \alpha_{n+1} \ln \zeta_{k+1}, \quad (5.6)$$

where $\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ is always the dominant solution in S_k .

Our next task is to rewrite (5.6) taking account of the dependence of a_j on the sequence of b 's and c 's. Let us begin with the case $n=3$.

$$\begin{aligned} (-1)^k \delta s_k &= dc_2 \left(\int_{\infty_k}^{\zeta_{n+1}} (\zeta \psi_1^2 + i b_2 \psi_1 \psi_2) d\zeta - i s_k b_3 \ln \zeta_{k+1} \right) \\ &+ db_2 \left(- \int_{\infty_k}^{\zeta_{k+1}} (\zeta \psi_2^2 - i c_2 \psi_1 \psi_2) d\zeta - i s_k c_3 \ln \zeta_{k+1} \right) \\ &+ dc_3 \left(\int_{\infty_k}^{\zeta_{k+1}} \psi_1^2 d\zeta - i s_k b_2 \ln \zeta_{k+1} \right) \end{aligned}$$

$$+db_3 \left(- \int_{\infty_k}^{\zeta_{k+1}} \psi_2^2 d\zeta - i s_k c_2 \ln \zeta_{k+1} \right) \quad (5.7)$$

Define

$$V_k = \begin{pmatrix} \int_{\infty_k}^{\zeta_{k+1}} \psi_2^2 d\zeta + i s_k c_2 \ln \zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} \psi_1^2 d\zeta - i s_k b_2 \ln \zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{n+1}} (\zeta \psi_2^2 - i c_2 \psi_1 \psi_2) d\zeta + i s_k c_2 \ln \zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} (\zeta \psi_2^2 + i b_2 \psi_1 \psi_2) d\zeta - i s_k b_3 \ln \zeta_{k+1} \end{pmatrix}. \quad (5.8)$$

Defining J as in (3.31) and $dR = (d\bar{c}_2, d\bar{b}_2, \dots, d\bar{c}_n, d\bar{b}_n)^T$, where (\bar{c}_j, \bar{b}_j) are the conjugate variables of Section 3, we have

$$(-1)^k s_k = \langle dR, J V_k \rangle, \quad k \in \mathbb{Z}, \quad (5.9)$$

where $\langle u, v \rangle = \sum_{i=1}^n u_i v_i$. By rewriting the identity

$$\int_{\zeta_1}^{\zeta_2} (b\psi_2^2 + c\psi_1^2) \frac{d\zeta}{\zeta} = (\psi_1 \psi_2) \Big|_{\zeta_1}^{\zeta_2} = -s_1 \quad (5.10)$$

in terms of (\bar{b}_k, \bar{c}_k) , we find

$$-s_k = \langle FR, J V_k \rangle, \quad k \in \mathbb{Z} \quad (5.11)$$

where F is the matrix $\{(-1)^{i+1} \delta_{ij}\}$. In Appendix 2, we find the orthogonality relations for the sequence of vectors $V = \{V_k\}$, $k \in \mathbb{Z}$,

$$M_k = \langle J V_k, V_\ell \rangle = \pi(\delta_{k, \ell-1} + s_k s_\ell), \quad k < \ell, \quad (5.12)$$

The set V forms a basis in \mathbb{R}^{2n-2} and the relations (5.9), (5.11) and (5.12) allow us to expand dR and FR in the basis V . Let $N = M^{-1}$; then

$$dR = \sum_{j, \ell} (-1)^j N_{\ell j} \delta s_j V_\ell \quad (5.13)$$

and

$$FR = \sum_{j, \ell} N_{\ell j} s_j V_\ell \quad (5.14)$$

where in both cases, the summation is over $j, l \in \mathbb{Z}$.

Several remarks are now in order.

1. We can, by the Gram-Schmidt process, choose a new set of vectors $U=U_k$, $k \in \mathbb{Z}$, such that $\langle JU_k, U_l \rangle = \delta_{k, l-1}$, $k \leq l$. The transformation between U and V will depend on the Stokes multipliers. We will not do this here but will indicate, in the results which follow, the effects of such a transformation.
2. Equation (5.9) shows us that if R changes in such a way such that $dR=0$ then $\delta s_k=0$. But $dR=0$ implies

$$\delta R = -J \nabla H_0 \delta t_0 - J \nabla H_1 \delta x - \sum_{k=2}^{n-2} J \nabla H_k \delta t_k$$

that is, R changes as a linear combination of the flows generated by Ω . Conversely if $\delta s_k=0$, then $dR=0$.

3. Notice that if

$$dR = -\alpha FR$$

then $\delta s_k = (-1)^k \alpha s_k$. However, the right hand side still is just the scaling flow and is proportional to $J \nabla H_0$. The time t dependence ($\delta = d/dt$ say) of s_k ($= s_k(t=0)e^{(-1)^k \alpha t}$).

can always be removed by correctly normalizing the asymptotic expansion (4.4).

4. Equation (5.9) is the starting point for a perturbation theory. Using it we may calculate how a perturbation

$$dR = \epsilon F(R), \quad 0 < \epsilon \ll 1, \quad (5.15)$$

affects the Stokes multipliers. In particular, (5.15) may no longer have the special property that as functions of the phases, the only moving singularities of q, r are poles. How is this reflected in the change of s_j ? More generally, it is very natural to ask if an analogue to the KAM (Kolmogoroff, Arnold, Moser) theorem obtains. The finite dimensional solution manifold for these flows is not necessarily compact, is not a torus and so the KAM theorem does not directly apply. The potential connection between a possible preservation of the solution manifold and the preservation of the Painlevé property is an intriguing one. Also, what would be the analogue of the tori which are not preserved? Are there resonances? The only frequencies which appear in the unperturbed problem are constants, either zero or one.

We now return briefly to some further details in the derivation of (5.9). Why, in general, does the expression

$$E = 1/\zeta (2da\psi_1\psi_2 - db\psi_2^2 + dc\psi_1^2)$$

naturally involve the conjugate variables $d\bar{c}_j$, $d\bar{b}_j$? The answer is that the basis vectors V_k must be carefully constructed in order that their inner products can be calculated. For example, when $n=4$, ignoring the terms arising from δt_j ,

$$E = dc_2(\zeta^2\psi_2^2 + ib_2\zeta\psi_1\psi_2 + ib_3\psi_1\psi_2) - db_2(\zeta^2\psi_1^2 - ic_2\zeta\psi_1\psi_2 - ic_3\psi_1\psi_2)$$

$$+ dc_3(\zeta\psi_1^2 + ib_2\psi_1\psi_2) - db_3(\zeta\psi_2^2 - ic_2\psi_1\psi_2) + dc_4\psi_1^2 - db_4\psi_2^2. \quad \text{If one were to define } V_k \text{ by}$$

writing E as an inner product between $(dc_2, db_2, dc_3, db_3, dc_4, db_4)^T$ and JV_k , the formula (A.14) of Appendix 2 would not apply. It turns out we must add

$-\frac{1}{2}a_3\psi_1^2 - \frac{1}{8}b_2^2\psi_2^2$ and $-\frac{1}{2}a_3\psi_2^2 - \frac{1}{8}c_2^2\psi_1^2$ to the coefficients of dc_2 and $-db_2$. Subtracting these terms from the rest of the expression E leads to a redefinition of b_4 and c_4 to \bar{b}_4 and \bar{c}_4 . Also there are terms proportional to δt_j left over which combine in just the correct way to make (5.9) hold. It now also becomes clear why α_{n+1} must factor into a sum of products of the conjugate variables. These terms cancel the potential logarithmic divergences. For $n=4$, the basis vectors are

$$\begin{aligned}
 & \int_{\infty_k}^{\zeta_{k+1}} d\zeta \{ (\zeta^2 - ia_3/2) \psi_2^2 - ic_2 \psi_1 \psi_2 - ic_3 \psi_1 \psi_2 - \frac{1}{8} c_2^2 \psi_1^2 \} - is_k \bar{c}_4 \ell n \zeta_{k+1} \\
 & \int_{\infty_k}^{\zeta_{k+1}} d\zeta \{ (\zeta^2 - \frac{ia_3}{2}) \psi_1^2 + ib_2 \psi_1 \psi_2 + ib_3 \psi_1 \psi_2 - \frac{1}{8} b_2^2 \psi_2^2 \} - is_k \bar{b}_4 \ell n \zeta_{k+1} \\
 & \int_{\infty_k}^{\zeta_{k+1}} d\zeta \{ (\psi_2^2 \theta - ic_2 \psi_1 \psi_2) \} - is_k \bar{c}_3 \ell n \zeta_{k+1} \\
 v_k = & \int_{\infty_k}^{\zeta_{k+1}} d\zeta \{ \zeta \psi_1^2 + ib_2 \psi_1 \psi_2 \} - is_k \bar{b}_3 \ell n \zeta_{k+1} \\
 & \int_{\infty_k}^{\zeta_{k+1}} d\zeta \psi_2^2 - is_k \bar{c}_2 \ell n \zeta_{k+1} \\
 & \int_{\infty_k}^{\zeta_{k+1}} d\zeta \psi_1^2 - is_k \bar{b}_2 \ell n \zeta_{k+1}.
 \end{aligned} \tag{5.16}$$

Recall that the vector $\psi = (\psi_1, \psi_2)^T$ is the dominant solution in sector S_k .

5b. Poisson brackets of the Stokes multipliers.

We have already defined the Poisson bracket in (3.29). Since δs_k does not depend on δX , $\delta T_2, \dots, \delta T_{n-2}$, we have

$$\begin{aligned}
 [s_k, s_\ell] &= \frac{\partial}{\partial \bar{b}_j} \frac{\partial s_k}{\partial \bar{c}_{n+2-j}} - \frac{\partial s_k}{\partial \bar{c}_{n+2-j}} \frac{\partial s_\ell}{\partial \bar{b}_j} \\
 &= -\langle \nabla s_k, J \nabla s_\ell \rangle \\
 &= -\langle (-1)^k J V_k, (-1)^\ell J^2 V_\ell \rangle \text{ from (5.9),} \\
 &= (-1)^{k+\ell} M_{k\ell}.
 \end{aligned} \tag{5.17}$$

Then the (k, ℓ) element of $P = \{[s_k, s_\ell]\}$, the matrix of Poisson brackets, is $(-1)^{k+\ell} M_{k\ell}^{-1}$. Define $N = M^{-1}$; then the (k, ℓ) element of $(P^{-1})^T$ is $(-1)^{k+\ell} N_{\ell k}$.

5c. Preservation of the symplectic form.

We now calculate the symplectic form

$$\begin{aligned} \omega &= \sum_j^n \delta \bar{b}_j \wedge \delta \bar{c}_{n+2-j} + \delta x \wedge \delta X + \sum_j^{n-2} \delta t_j \wedge \delta T_j \\ &= -\langle \delta_1 R, J \delta_2 R \rangle + \delta x \wedge \delta X + \sum_j^{n-2} \delta t_j \wedge \delta T_j \end{aligned} \quad (5.18)$$

in terms of the infinitesimal variations of the new coordinates $s_k, \delta x, \delta t_j, \delta H_1, \delta H_j$. First, recall that

$$\delta R = dR - J \nabla H_0 \delta t_0 - J \nabla H_1 \delta x - \sum_k^{n-2} J \nabla H_k \delta t_k. \quad (5.19)$$

Second, observe that since

$$\begin{aligned} X &= H_1 - \hat{H}_1(\bar{b}_k, \bar{c}_k, x, t_k), \quad T_j = H_j - \hat{H}_j(\bar{b}_k, \bar{c}_k, x, t_k), \text{ we have} \\ \delta X &= \delta H_1 - \langle \nabla H_1, \delta R \rangle - H_{1x} \delta x - \sum_k^{n-2} H_{1t_k} \delta t_k \\ &= \delta H_1 - \langle \nabla H_1, dR \rangle + (\langle \nabla H_1, J \nabla H_1 \rangle - H_{1x}) \delta x \\ &\quad + \sum_k^{n-2} (\langle \nabla H_1, J \nabla H_k \rangle - H_{1t_k}) \delta t_k, \end{aligned} \quad (5.20)$$

$$\begin{aligned} \delta T_j &= \delta H_j - \langle \nabla H_j, dR \rangle + (\langle \nabla H_j, J \nabla H_1 \rangle - H_{jx}) \delta x \\ &\quad + \sum_k^{n-2} (\langle \nabla H_j, J \nabla H_k \rangle - H_{jt_k}) \delta t_k. \end{aligned} \quad (5.21)$$

In (5.20), (5.21) H_{1x} means $\frac{\partial H_1}{\partial x}$, the partial derivative of H_1 with respect to explicit x . Also in both $\delta X, \delta T_j$, the terms proportional to δt_0 are of the form $\langle J \nabla H_0, \nabla H_k \rangle = [H_k, H_0] = 0$ (since H_0 is independent of t_k and t_0 never explicitly appears anywhere). In these equations we have also replaced $\nabla \hat{H}_k, \hat{H}_{kx}$ by ∇H_k and H_{kx} since these are respectively equal. Also we recall that since $J^T = -J$, $\langle Ju, v \rangle = \langle v, Ju \rangle = -\langle u, Jv \rangle$. After a little calculation we now find,

$$\omega = \langle -d_1 R, J d_2 R \rangle + \delta x \wedge \delta H_1 + \sum_j^{n-2} \delta t_j \wedge \delta H_j. \quad (5.22)$$

The terms proportional to $\delta x \wedge \delta t_j$ and $\delta t_j \wedge \delta t_k$ have as coefficients $[H_j, H_1]$ and $[H_k, H_j]$ respectively, which are zero.

But from (5.13),

$$\begin{aligned}
 -\langle d_1 R, J d_2 R \rangle &= -\langle d_1 R, j_{\Sigma, \ell} (-1)^{j+\ell} N_{j\ell} \delta_2 s_j V_\ell \rangle \\
 &= j_{\Sigma, \ell \in \mathbb{Z}} (-1)^{j+\ell} N_{j\ell} \delta_1 s_\ell \delta_2 s_j \text{ from (5.9),} \\
 &= (\delta_1 s)^T (P^{-1})^T \delta_2 s \\
 &= \langle P^{-1} \delta_1 s, \delta_2 s \rangle
 \end{aligned}$$

where P is the matrix of Poisson brackets. By a judicious choice of basis U such that $\langle JU_k, U_\ell \rangle = \delta_{k, \ell-1}$, $k \leq \ell$, we can arrange combinations $f_j(s_k)$, $g_j(s_k)$ $j \in \mathbb{Z}$ such that

$$\omega = j_{\Sigma} \delta f_j(s_k) \wedge \delta g_j(s_k) + \delta x \wedge \delta H_1 + \sum_{j=2}^{n-2} \delta t_j \wedge \delta H_j. \quad (5.23)$$

In (5.23), the only variables which depend on x, t_j are the variables x, t_j $j=2, \dots, n-2$ themselves. For the t_k flow ($t_1=x$) $f_j, g_j, H_1, H_k, k=2, \dots, n-2$ are constant and

$$t_k = \frac{\partial H_k}{\partial H_1} = 1.$$

5d. Fourier expansions and contour integral representations.

We focus on the case $n=3$ and take $t_0=0$. The equations (3.13) are

$$q_{xx} - 2q^2 r = xq, r_{xx} - 2qr^2 = rxr. \quad (5.24)$$

From (5.14),

$$FR = -\sum_{j, \ell} N_{j\ell} s_j V_\ell \quad (5.25)$$

with V given by (5.8). In this case, there are six sectors at $\zeta=\infty$. We will choose V as being formed from sectors 1, 2, 4 and 5 and the relevant Stokes multipliers are s_1, s_2, s_4, s_5 . In this case

$$N = \frac{1}{\pi(1+s_1 s_2)(1+s_4 s_5)} \begin{pmatrix} 0 & -1-s_4 s_5 & s_2 s_5 & -s_2 s_4 \\ 1+s_4 s_5 & 0 & -s_1 s_5 & s_1 s_4 \\ -s_2 s_5 & s_1 s_5 & 0 & -1-s_1 s_2 \\ s_2 s_4 & -s_1 s_4 & 1+s_1 s_2 & 0 \end{pmatrix}$$

and ($\alpha_1=i$)

$$\pi \begin{pmatrix} c_2 \\ -b_2 \\ c_3 \\ -b_3 \end{pmatrix} = \pi \begin{pmatrix} r \\ -q \\ \frac{-ir_x}{2} \\ \frac{-iq_x}{2} \end{pmatrix} = \frac{s_2}{1+s_1s_2} V_1 - \frac{s_1}{1+s_1s_2} V_2 + \frac{s_5}{1+s_4s_5} V_4 - \frac{s_4}{1+s_4s_5} V_5. \quad (5.26)$$

Case 1. The connection with the squared eigenfunction expansions of inverse scattering.

Originally we were looking for a four parameter (complex) family of solutions for (5.24). Let us look at the two parameter family formed by setting $s_2=s_5=0$. Then, from (5.26)

$$\begin{pmatrix} q \\ -r \end{pmatrix} = \frac{s_1}{\pi} \int_2^{\infty} \begin{pmatrix} \psi_1^2 \\ \psi_2^2 \end{pmatrix} d\zeta + \frac{s_4}{\pi} \int_4^{\infty} \begin{pmatrix} \psi_1^2 \\ \psi_2^2 \end{pmatrix} d\zeta \quad (5.27)$$

where ψ refers to the uniquely defined dominant solution in each of the Sectors 2 and 5. But the ψ in S_2 is simply the recessive solution in S_1 and, since $s_2=0$, also the recessive solution in S_3 . Hence $\psi e^{-i\zeta^3/3}$ is a solution of both (3.1a) and

(3.1c) which is analytic for $\text{Im}\zeta > 0$ and asymptotes to $\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x}$ as $\zeta \rightarrow \infty$ in $\text{Im}\zeta > 0$. This is precisely the solution ψ we defined in papers [9,10] in connection with the scat-

tering problem (3.1a). Similarly the solution $\psi e^{i\zeta^3/3}$ of S_5 is what we called $\bar{\psi}$ in [9] and [10]. Now since the ψ of $S_2(S_5)$ is recessive in $S_1, S_3(S_4, S_6)$, we may ex-

tend the end points of the integration paths in (5.32) to $-\infty$ and ∞ on the real ζ axis. Thus (5.27) becomes

$$\begin{pmatrix} q \\ -r \end{pmatrix} = \frac{-1}{\pi} \int_{-\infty}^{\infty} s_1 e^{\frac{2}{3}\zeta^3} \begin{pmatrix} \psi_1^2 \\ \psi_2^2 \end{pmatrix} d\zeta + \frac{1}{\pi} \int_{-\infty}^{\infty} s_4 e^{-\frac{2}{3}\zeta^3} \begin{pmatrix} \bar{\psi}_1^2 \\ \bar{\psi}_2^2 \end{pmatrix} d\zeta \quad (5.28)$$

which is precisely equation 6.55 of reference [10] with

$$\frac{b}{a}(\zeta) = +s_1 e^{\frac{2}{3}\zeta^3} \quad \frac{\bar{b}}{\bar{a}} = s_4 e^{-\frac{2}{3}\zeta^3} \quad (5.29)$$

But if we do define $b/a, \bar{b}/\bar{a}$ in the usual way for (3.1a), then (3.1c) shows that

$$\frac{b}{a}(\zeta) = \frac{b}{a}(0) e^{\frac{2}{3}\zeta^3}, \quad \frac{\bar{b}}{\bar{a}}(\zeta) = \frac{\bar{b}}{\bar{a}}(0) e^{-\frac{2}{3}\zeta^3} \quad (5.30)$$

and so the Stokes multipliers s_1 and s_4 are simply $b/a(\zeta=0)$ and $-\bar{b}/\bar{a}(\zeta=0)$ respectively. We showed in [1], that if $r=q, s_4=s_1$. We also know in this case that

$\bar{b}(\zeta) = -b(-\zeta)$ and $\bar{a}(\zeta) = a(-\zeta)$ which is consistent.

Case 2. Contour integral representations of Painlevé functions.

Again we let $r=q$. From previous work [1] we know $s_4=s_1$, $s_5=s_2$ and the ψ of $S_4(S_5)$ is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tilde{\psi}(-\zeta)$ where the latter refers to the recessive solution of sector 1 (2). Changing the integration variable in V_4 and V_5 of (5.26), we obtain

$$\begin{aligned} q &= \frac{s_2}{1+s_1 s_2} \int_{\infty_1}^{\zeta_2} (\psi_1^2 - \psi_2^2) d\zeta - 2is_1 q \ln \zeta_2 \\ &= \frac{s_1}{1+s_1 s_2} \int_{\infty_2}^{\zeta_3} (\tilde{\psi}_1^2 - \tilde{\psi}_2^2) d\zeta - 2is_2 q \ln \zeta_3 \end{aligned} \quad (5.31)$$

where $\psi, (\tilde{\psi})$ are the dominant (recessive) solutions of S_1 . Now the contour of the first integral can be extended back to ∞_6 as ψ is recessive in S_6 . Similarly, we can set the ∞_2 of the second integral to ∞_1 . Now the contours are the same as those used in the integral definitions of Airy functions. In particular, if $s_2=0$ which we have shown is equivalent to picking that class of solutions of (5.24) which decay at $x=\pm\infty$ and admit construction by inverse scattering,

$$q = +s_1 \int_{-\infty}^{\infty} (\tilde{\psi}_1^2 - \tilde{\psi}_2^2) d\zeta \quad (5.32)$$

In the limit of small amplitudes,

$$\tilde{\psi} \sim \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta^3/3 + i\zeta x}$$

and (5.) is

$$q = -s_1 \int_{-\infty}^{\infty} e^{2i\zeta x + 2i\zeta^3/3} d\zeta$$

which is $-s_1 \pi A_1(x/4)$.

These representations will be very important when we treat the equivalence between Painlevé equations and particle systems which parallels the analogy between the finite gap solutions and particle systems [5].

Appendix I: Asymptotic Expansions.

We write equation (3.1c) in the form

$$V_\zeta = \frac{1}{\zeta} (-aY_1 + bY_2 + cY_3) V \quad (A.1)$$

in the notation (2.3). Let

$$V \sim \left(I + \sum_{r=1}^{\infty} \frac{C_r}{\zeta^r} \right) e^{-\theta Y_1 - \psi} \quad (A.2)$$

with

$$\theta_\zeta = \alpha_1 \zeta^{n-1} + \alpha_3 \zeta^{n-3} + \dots + (\alpha_n + ix) + \frac{1}{\zeta} \alpha_{n+1} + \dots \quad (A.3)$$

$$\psi_3 = \frac{\delta_{n+3}}{z^3} + \dots \quad (\text{A.4})$$

In (A.3), the integration constant is it_0 , in (A.4) it is zero. By induction, we show that for $k \leq n$

$$c_k = -\frac{1}{2}(\delta_{k+1}y_2 - \hat{c}_{k+1}y_3) \quad (\text{A.5})$$

where δ_{k+1} , \hat{c}_{k+1} are defined recursively as

$$\delta_{k+1} = \delta_{k+1} + \frac{1}{2} \sum_{\ell=3}^k (a_\ell + \alpha_\ell) \delta_{k-\ell+2}, \hat{c}_{k+1} = c_{k+1} + \frac{1}{2} \sum_{\ell=3}^k (a_\ell + \alpha_\ell) \hat{c}_{k-\ell+2} \quad (\text{A.6})$$

$$\delta_2 = \delta_2, \delta_3 = \delta_3, c_2 = \hat{c}_2, \hat{c}_3 = c_3.$$

Also

$$\hat{\alpha}_{n+1-j} = \alpha_{n+1-j} \quad j=2, \dots, n. \quad (\text{A.7})$$

After this stage, the terms from differentiating the bracket term on the right hand side of (A.2) enters and disrupts the pattern.

$$c_{n+1} = -\frac{1}{2}(\delta_{n+2} - \frac{1}{2}\delta_2)y_2 + \frac{1}{2}(\hat{c}_{n+2} + \frac{1}{2}\hat{c}_2)y_3,$$

$$c_{n+2} = -\frac{1}{2}(\delta_{n+3} - i\delta_3)y_2 + \frac{1}{2}(\hat{c}_{n+3} + i\hat{c}_3)y_3,$$

$$c_{n+3} = -\frac{1}{2}(\delta_{n+4} - \frac{3i}{2}\delta_4 - \frac{1}{2}\delta_{n+3}\delta_2)y_2 \\ + \frac{1}{2}(\hat{c}_{n+4} + \frac{3i}{2}\hat{c}_4 + \frac{1}{2}\delta_{n+3}\hat{c}_2)y_3,$$

$$\text{with } \delta_{n+3} = \frac{b_2c_2}{4}, \delta_{n+4} = 3/8(b_2c_3 + b_3c_2),$$

$$\hat{\alpha}_{n+4} = \alpha_{n+4} + 1/8(b_2c_3 - b_3c_2).$$

In particular we often have occasion to need to calculate $\psi_1\psi_2$ on R_1 and R_2 as

$z \rightarrow \infty$. On R_1 , $\psi \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e$, and thus $\psi_1\psi_2 \sim 0$. On R_2 , however, $\psi \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^\theta - s_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-\theta}$ and thus

$\psi_1\psi_2 \sim -s_1$. In order to calculate the behavior of the components of V_k , we need to keep more terms in the expansions.

Appendix II: The inner product (5.12).

We will illustrate the means of proof in the case $n=3$. Let us examine

$$\langle JV_1, V_2 \rangle = \left\{ \begin{array}{l} \int_{\zeta_1}^{\zeta_2} (\zeta \psi_1^2 + i b_2 \psi_1 \psi_2) d\zeta - i s_1 b_3 \ln \zeta_2 \\ - \int_{\zeta_1}^{\zeta_2} (\zeta \psi_2^2 - i c_2 \psi_1 \psi_2) d\zeta - i s_1 c_3 \ln \zeta_2 \\ \int_{\zeta_1}^{\zeta_2} \psi_1^2 d\zeta - i s_1 b_2 \ln \zeta_2 \\ - \int_{\zeta_1}^{\zeta_2} \psi_2^2 d\zeta - i s_1 b_2 \ln \zeta_2 \end{array} \right\} \cdot \left\{ \begin{array}{l} \int_{\zeta_2}^{\zeta_3^+} \bar{\psi}_2^2 d\zeta + i s_2 c_2 \ln \zeta_3^+ \\ \int_{\zeta_3}^{\zeta_3^+} \bar{\psi}_1^2 d\zeta - i s_2 b_2 \ln \zeta_3^+ \\ \int_{\zeta_2}^{\zeta_3} (\zeta \bar{\psi}_2^2 - i c_2 \bar{\psi}_1 \bar{\psi}_2) d\zeta + i s_2 c_3 \ln \zeta_3 \\ \int_{\zeta_2}^{\zeta_3^+} (\zeta \bar{\psi}_1^2 + i b_2 \bar{\psi}_1 \bar{\psi}_2) d\zeta - i s_2 b_3 \ln \zeta_3^+ \end{array} \right\} \quad (A.9)$$

ψ is the dominant solution of S_1 and $\bar{\psi}$, the recessive solution of S_1 , is the dominant solution of S_2 . We take ζ_2^+ to be closer to $\zeta = \infty$ than ζ_2 on R_2 . The product in (A.9) has three sets of terms. First, the terms in products of logarithms cancel.

Next, the terms proportional to $+i s_2 \log \zeta_3^+$ are

$$\int_{\zeta_1}^{\zeta_2} (c_2 \zeta + c_3) \psi_1^2 + (b_2 \zeta + b_3) \psi_2^2 d\zeta = (\psi_1 \psi_2) \frac{\zeta_2}{\zeta_1}$$

which in the limit of ζ_1, ζ_2 tending to $\zeta = \infty$ on R_1, R_2 respectively is $-s_1$ (use Appendix I). Similarly the terms proportional to $-i s_1 \ln \zeta_2$ are $-s_2$. Hence these two terms together contribute $-i s_1 s_2 \ln \zeta_3^+ / \zeta_2$. Next we examine the integral terms which may be written

$$\int_{\zeta_2}^{\zeta_3^+} \int_{\zeta_1}^{\zeta_2} d\zeta d\zeta' \{ (\zeta + \zeta') (\psi_1^2 \bar{\psi}_2^2 - \bar{\psi}_1^2 \psi_2^2) - i b_2 (\psi_2^2 \bar{\psi}_1 \bar{\psi}_2 - \bar{\psi}_2^2 \psi_1 \psi_2) - i c_2 (\psi_1^2 \bar{\psi}_1 \bar{\psi}_2 - \bar{\psi}_1^2 \psi_1 \psi_2) \} \quad (A.11)$$

where ζ lies in the path between ζ_1, ζ_2 and ζ' on the path between ζ_2^+, ζ_3^+ . Now the integral is

$$-i \left(\frac{a}{\zeta} - \frac{a'}{\zeta'} \right) (\psi_1^2 \bar{\psi}_2^2 - \bar{\psi}_1^2 \psi_2^2) + \left(\frac{b}{\zeta} - \frac{b'}{\zeta'} \right) (\psi_2^2 \bar{\psi}_1 \bar{\psi}_2 - \bar{\psi}_2^2 \psi_1 \psi_2) + \left(\frac{c}{\zeta} - \frac{c'}{\zeta'} \right) (\psi_1^2 \bar{\psi}_1 \bar{\psi}_2 - \bar{\psi}_1^2 \psi_1 \psi_2) \quad (A.12)$$

We have an identity: If A,B,C satisfy

$$A_{\zeta} = BC + \gamma B, \quad B_{\zeta} + 2\alpha B = 2\beta A, \quad C_{\zeta} - 2\alpha C = 2\gamma A \quad (\text{A.13})$$

and A', B', C' satisfy the same equations with ζ replaced by ζ' , then

$$\begin{aligned} & (\alpha - \alpha')(BC' - B'C) + (\beta - \beta')(A'C - AC') + (\gamma - \gamma')(A'B - AB') \\ &= \frac{1}{2} \left(\frac{d}{d\zeta} + \frac{d}{d\zeta'} \right) (2AA' - B'C - BC'). \end{aligned} \quad (\text{A.14})$$

Now, in our case $A = \psi_1 \psi_2$, $B = \psi_1^2$, $C = \psi_2^2$, $\alpha = \frac{a}{\zeta}$, $\beta = \frac{b}{\zeta}$, $\gamma = \frac{c}{\zeta}$ and thus (A.11) is

$$-\frac{i}{2} \int_{\zeta_3^+}^{\zeta_2^+} \int_{\zeta_1^+}^{\zeta_2} d\zeta d\zeta' \frac{d}{d\zeta} + \frac{d}{d\zeta'} \frac{W(\zeta, \zeta')}{\zeta - \zeta'} \quad (\text{A.15})$$

where

$$W(\zeta, \zeta') = -(\psi_1(\zeta) \bar{\psi}_2(\zeta') - \psi_2(\zeta) \bar{\psi}_1(\zeta'))^2.$$

Integrating we obtain (A.11) is

$$-\frac{1}{2} \int_{\zeta_1^+}^{\zeta_2} \frac{W(\zeta, \zeta_3^+)}{\zeta - \zeta_3^+} - \frac{W(\zeta, \zeta_2^+)}{\zeta - \zeta_2^+} d\zeta - \frac{1}{2} \int_{\zeta_2^+}^{\zeta_3^+} \frac{W(\zeta_2, \zeta')}{\zeta_2 - \zeta'} - \frac{W(\zeta_1, \zeta')}{\zeta_1 - \zeta'} d\zeta'. \quad (\text{A.16})$$

Now, we take the path between ζ_1 and ζ_2 to be in along R_1 to the $\zeta=0$ and out along R_2 . To integrate the third integral, we recognize that $\bar{\psi}$ is recessive in S_1 and bring the contour from ζ_2^+ from $\text{Arg } \zeta = \frac{\pi}{3}$ to $\text{Arg } \zeta = 0$, then in along R_1 and out R_3 to ζ^+ . This means we must take account of the pole at $\zeta' = S_2$ and since

$$\text{Res. } \frac{W(\zeta_2, \zeta')}{\zeta_2 - \zeta'} = -W(\zeta_2, \zeta_2) = -1, \text{ since } \psi_1 \bar{\psi}_2 - \bar{\psi}_1 \psi_2 \text{ is the Wronskian. Hence from}$$

the third integral we obtain the extra term $-1/2(-2\pi i)(-1) = \pi$. Now on the neutral rays, the only terms which will contribute as $\zeta_1, \zeta_2, \zeta_2^+, \zeta_3^+$ tend to infinity are those along rays where $W(\zeta, \zeta')$ tends to a constant. This only happens when ζ lies on R_2 , ζ' on R_3 whence $W(\zeta, \zeta') \rightarrow +2s_1 s_2$, and thus the remaining terms in (A.16) tend asymptotically to

$$\pi - is_1 s_2 \int_0^{\zeta_2} \frac{d\zeta}{\zeta - \zeta_3^+} - is_1 s_2 \int_0^{\zeta_3^+} \frac{d\zeta'}{\zeta_2 - \zeta'}.$$

Hence

$$\langle JV_1, V_2 \rangle = \pi - is_1 s_2 \ln \frac{\zeta_2 - \zeta_3^+}{-\zeta_3^+} + is_1 s_2 \ln \frac{\zeta_3^+ - \zeta_2}{-\zeta_2} - is_1 s_2 \ln \frac{\zeta_3^2}{\zeta_2} = \pi(1 + s_1 s_2).$$

It is easy to see that if S_k and S_ℓ ($k \neq \ell$) are not contiguous sectors, the contribution from the pole disappears. Hence we have

$$\langle JV_k, V_\ell \rangle = (\pi \delta_{k, \ell-1} + s_k s_\ell) \quad (\text{A.17})$$

This formula also holds for all $n \geq 3$.

Acknowledgements:

The authors gratefully acknowledge support for this work Grant #DAAG29-81-K-0025, N00014-76-C-0867 and MCS-7903498-01.

References

- [1] Flaschka, H. and Newell, A.C., Monodromy and Spectrum Preserving Deformation I, Comm. Math. Phys. 76 (1980) 65-116.
- [2] Flaschka, H. and Newell, A.C., Multiphase Similarity Solutions of Integrable Evolution Equations, Lectures in Pure and Applied Mathematics 54 (1980) 373-395.
Also, Proc. Kiev Conference on Solitons, September 1979, to appear Physica D July, 1981.
- [3] Lax, P.D., Periodic Solutions of the Korteweg deVries equation, Lectures in Applied Math. 15 (1974) 85-96.
- [4] Novikov, S.P., The Periodic Problem for the Korteweg deVries Equation, Funkts. Anal. Prilozhen 8 (1974) 54-66.
- [5] Flaschka, H. and Newell, A.C., Monodromy and Spectrum Preserving Deformations II, III. In preparation.
- [6] Newell, A.C., The General Structure of Integrable Evolution Equations, Proc. Roy. Soc. A 365, (1979) 283-311.
- [7] Jimbo, M., Miwa, T., Mori, Y. and Sato, J., Density Matrix of an Inpenetrable Bose Gas and the Fifth Painlevé Transcendent, Physica D 1, (1980) 80-158.
- [8] Gel'fand, I.M. and Dikii, L.A., Integrable Nonlinear Equations and the Liouville Theorem, Funkt. Analiz Priloz. 1, (1979) 8-20.
- [9] Ablowitz, M.J., Kaup, D.J., Newell, A.C. and Segur, H., The Inverse Scattering Transform, Stud. Appl. Math. 53, (1974) 249.
- [10] Newell, A.C., The Inverse Scattering Transform, Topics in Current Physics 17, (1980) 177-242. Solitons, Eds. R. Bullough and P. Caudrey, Springer-Verlag.

This Page Intentionally Left Blank

NUMERICAL METHODS FOR NONLINEAR DIFFERENTIAL EQUATIONS

James M. Hyman

Center for Nonlinear Studies
Theoretical Division, MS 610
Los Alamos National Laboratory
Los Alamos, New Mexico 87544
U.S.A.

New and better methods for the numerical solution of partial differential equations are being developed at an ever-increasing rate. In this paper, directed to scientists trained in mathematics but not necessarily in numerical analysis, we try to unify and simplify the underlying crucial points in this development. Most of the new methods can be understood and classified according to how space, time, and boundary conditions are discretized and by how nonlinear algebraic equations that arise in the solution process are solved. We will discuss each point and present numerical examples showing how a simple linear analysis can fail.

INTRODUCTION

A numerical algorithm for the solution of nonlinear partial differential equations (PDEs) can be a highly complicated and problem-dependent process. A method developed for a particular test problem may not work for similar problems. Methods that work well in one space dimension may not be easily extended to two or three dimensions. Linear analysis can rarely ensure accuracy with nonlinear methods or highly nonlinear equations.

Each year significant and powerful algorithms are discovered, but most of these methods have a similar underlying structure. To better predict when a method, which is almost always developed for relatively simple test problems, will extend to more complicated situations, we must understand this underlying structure.

First, we will simplify the structure of these methods to study the general flow of the algorithms, understand their similar patterns of interconnections, and unify them in a single theory. In this synthesis, new properties and common features among seemingly different methods sometimes emerge that were not evident when analyzing a specific method for a specific set of equations.

We do not imply that methods tailored to specific equations should be abandoned. Many excellent methods have come from specialized analysis of a specific set of equations. But, by understanding the general patterns found in all the methods, we may gain a better view of how and why the algorithms work as they do.

The prototype system of PDEs studied here can be written as

$$u_t = f(x, t, u), \quad u(x, 0) = u_0, \quad (1)$$

where the solution $u(x, t)$ lies in some function space, x is in some domain Ω , and f is a nonlinear differential operator. We use the notation u_t and u_x to represent partial differentiation with respect to time and space.

On the boundary of Ω , the solution is constrained to satisfy the boundary condition

$$b(x,t,u(x,t)) = 0, \quad x \in \partial\Omega, \quad (2)$$

where b is a nonlinear spatial differential operator.

A discrete numerical method approximates u by an element U in some finite dimensional space whose components are the values of u at a discrete set of mesh points. The differential operators f and b are replaced by discrete operations F and B operating on U .

The discretized approximation to Eqs. (1) and (2) is a constrained system of ordinary differential equations (ODEs),

$$U_t = F(U), \quad B(U) = 0, \quad (3)$$

which are then integrated numerically.

Evaluating $F(U)$ and integrating Eq. (3) often requires solving large sparse systems of algebraic equations. The methods used to solve these equations often determine the success or failure of any numerical approach.

This paper is organized so that the crucial choices for methods to discretize space, boundary conditions, time, and methods to solve the algebraic systems are analyzed independently. The numerical examples are in only one space dimension for simplicity. We avoid taking undue advantage of this or any linearity that may be present so that the general conclusions may be extended to more complicated problems in higher dimensions.

SPACE DISCRETIZATION

The numerical approximation of the spatial derivatives and the distribution of the mesh points determine how well the spatial operator $f(u)$ and the solution u will be approximated. We describe some typical methods to approximate spatial derivatives and then we describe how the errors in a calculation are related to the order of accuracy of the method.

Often, important properties of the solution behavior originate at the boundary and the numerical differentiation procedure must take the boundary conditions into account. For simplicity, we will not incorporate the influence of the boundary conditions into the discrete operator until the next section.

Numerical Differentiation

The guiding principle in choosing a numerical method to approximate the spatial operator is that the resulting discrete model should retain as closely as possible all the crucial properties of the original differential operator. For instance, for a hyperbolic PDE, the operator f is antisymmetric, so we try to approximate f by an antisymmetric discrete operator F . For a parabolic PDE when f is dissipative, we approximate f with a dissipative discrete operator. If f is in conservation form, we also choose a conservation form of F .

All spatial differentiation methods we describe follow the same algorithmic flow. At time t during a calculation, we are given the approximate solution vector U at a discrete set of mesh points X and must generate a numerical approximation $F(U)$ of $f(u)$ at these mesh points. When $f(u)$ is a nonlinear spatial operator, it will have terms such as $g(u,x,t)_x$ or $\{s(u,x) g(u,x,t)\}_x$.

First, all nonlinear functions are evaluated to generate, say, the vectors G and S . Next, G , X , and S are input for a black-box-type subroutine package in the computer library. This subroutine then returns the approximate vectors G_x and $(SG_x)_x$ of derivatives at the mesh points as output.

Thus, the spatial differentiation is totally divorced from the nonlinearities of the PDE. This modularity also reduces the redundancy of programming the same approximation to the spatial derivatives each time they appear in an equation. These differentiation routines are debugged and optimized for a particular machine only once--with no specific PDE in mind.

The numerical differentiation methods we describe fall into three categories: finite difference methods, implicit methods, and transformation methods. The simplest and oldest general technique, the finite difference method, is described first.

Finite Differences

In a finite difference approximation, the function g is expanded at mesh points near x_i in a Taylor series about x_i . These expansions are added and subtracted to give

$$\frac{\partial g}{\partial x} \approx G_x(x_i) = \frac{G_{i+1} - G_{i-1}}{2\Delta x} - \frac{\Delta x^2}{6} g_{xxx}(x_i) + O(\Delta x^4) \quad (4)$$

$$\frac{\partial g}{\partial x} \approx G_x(x_i) = \frac{-G_{i+2} + 8G_{i+1} - 8G_{i-1} + G_{i-2}}{12\Delta x} + O(\Delta x^4) \quad (5)$$

$$\frac{\partial}{\partial x} S \frac{\partial g}{\partial x} \approx \frac{S_{i+\frac{1}{2}}(G_{i+1} - G_i) - S_{i-\frac{1}{2}}(G_i - G_{i-1})}{\Delta x^2} \quad (6)$$

where $\Delta x = \frac{1}{2}(x_{i+1} - x_{i-1})$ and

$$S_{i+\frac{1}{2}} = 2S(x_i)S(x_{i+1})/(S(x_i) + S(x_{i+1})) \quad (7)$$

Here $O(\Delta x^j)$ stands for a quantity bounded by some constant times Δx^j as Δx goes to zero. The formula for $(sg)_x$ uses the harmonic rather than the arithmetic average for $S_{i+\frac{1}{2}}$ to insure flux continuity in the solution [1]. This crucial property of the differential operator should be retained in the discrete approximation.

The vector G_x of derivative values can be written as $G_x = DG$. Here, D is a banded sparse matrix.

Implicit Methods

The implicit or Padé methods are a generalization of the finite difference methods. Equation (4) can be rewritten as

$$\left(1 + \frac{\Delta x^2}{6} \frac{\partial^2}{\partial x^2}\right) g_x(x_i) = \frac{G_{i+1} - G_{i-1}}{2\Delta x} + O(\Delta x^4) \quad (8)$$

Approximating the second derivative operation with Eq. (6) and using matrix notation, we have

$$D_2 G_x = D_1 G + O(\Delta x^4) \quad (9)$$

Here, D_2 and D_1 are tridiagonal matrices generated by Eqs. (4) and (6). A similar procedure can be used to approximate g_{xx} .

Often, time discretization methods require solving implicitly defined equations such as Eq. (9). In special cases the implicitly defined space and time systems can be solved simultaneously leading to an exceptionally high-order efficient method. These are called operator compact implicit methods [2].

Transformation Methods

In the transformation or pseudospectral method, G is first mapped by a transformation of the form

$$g = TG = \sum_{j=1}^m a_j \phi_j(x) \quad (10)$$

into nondiscrete function space. The basis functions $\phi_j(x)$ and the transformation are chosen so that T and T^{-1} are fast ($\leq m \log m$ operations), and so that differentiation D is simple in the transform space. The derivative approximation can then be written as

$$G_x = T^{-1}DTG \quad (11)$$

Some common transforms are based on the fast Fourier transform where the ϕ_j are trigonometric functions, Tchebyshev, or Legendre polynomials. Choosing the ϕ_j as piecewise polynomials with compact support, such as the B splines, is another good choice. Often the transformation is chosen to incorporate some crucial property such as the periodicity or symmetry of g into the calculation. This can greatly improve the accuracy of G_x .

Grid Resolution

The accuracy of the numerical solution is largely determined by how well $F(U)$ approximates $f(u)$ and how well U resolves the solution u . These errors depend on the accuracy of the derivative approximation and the algorithm used to locally refine the mesh. Because the accuracy of the method strongly influences the grid needed to resolve the solution, we first describe these relationships.

Accuracy

Discretizing with a highly accurate difference scheme keeps the computer time and storage to a reasonable level in multidimensional calculations. The analysis is linear, but in practice, qualitative results usually hold for nonlinear problems. Usually, the best we can do for nonlinear equations is to seek out basic relationships for linear equations that will be stable to small perturbations. These are the results that are most often retained.

The truncation error in (3) is the amount by which the mesh point values of the true solution to the differential equation (1) fail to satisfy the difference

equation (3). These errors are typically $O(\Delta x^j)$ in size and j is called the order of the method.

For any fixed accuracy criteria, the number of mesh points M_1 needed in a j th-order linear calculation is related to the number of mesh points needed by other methods according to the relationship

$$M_1 = C_2 M_2^2 = C_4 M_4^4 = C_6 M_6^6 . \quad (12)$$

For the periodic unidirectional wave equation $u_t = v u_x$, the phase error introduced will be the same using second-, fourth-, or sixth-order differences if the number of mesh points in the calculations satisfy [3]

$$M_2 \cong 0.36 M_4^2 \cong 0.12 M_6^3 . \quad (13)$$

Table 1 compares the number of points per wavelength necessary to obtain a given phase error e in the k th Fourier mode of the periodic solution to $u_t = v u_x$ at time t using second-, fourth-, and sixth-order spatial centered differences in (3).

2nd order M_2	4th order M_4	6th order M_6	Accuracy $e/(vkt)$
4	4	3	2.6
8	5	4	0.65
16	7	5	0.16
32	10	7	0.04
64	14	8	0.01
128	19	10	0.0025
256	27	13	0.0006

Table 1. Points per wavelength for second-, fourth-, and sixth-order differences to have the same accuracy.

In a calculation where the solution contains many different frequencies, the high modes (2-5 points per wavelength) are approximated equally poorly with all the methods. The middle modes (6-16 points per wavelength) are computed much more accurately with the fourth- and sixth-order differences than with the second-order method. The sixth-order differences are more accurate for the lower modes than either second- or fourth-order differences, but this gain is often lost because of errors introduced in the approximation of the boundary conditions.

The relationship of the accuracies of the different methods compared to the number of points per wavelength is even more impressive in higher dimensions. In two space dimensions, the number in Table 1 should be squared, and in three dimensions, cubed.

The corresponding relationship for the damping error in $U_t = U_{xx}$ for second- vs fourth-order finite differences is [3]

$$M_2 = 0.44 M_4^2 . \quad (14)$$

The qualitative results from this linear analysis hold true for many nonlinear equations, as shown in Fig. 1 where the density is plotted into solutions of

the Euler equations of gas dynamics for a Riemann problem. (A complete description of these calculations is in Ref. 4.) Second-order finite differences were used in Fig. 1a and fourth-order in Fig. 1b, otherwise they are the same. Note that the higher order differences resolve the solution better.

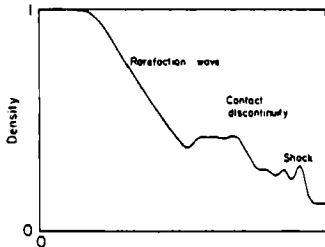


Figure 1a
Second-order

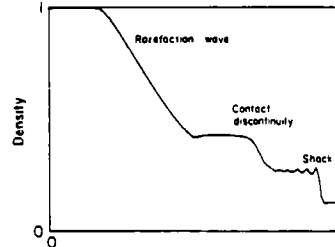


Figure 1b
Fourth-order

Figure 1

Numerical Solutions to the Euler Equations for the Riemann Problem
on a Grid of 50 Mesh Points

The oscillations near the shock are caused by errors in the high frequencies where all the methods do poorly. To efficiently resolve the gradients in this problem, it would be best to adaptively refine the mesh around the discontinuities [5]. These methods either continuously redistribute a fixed number of mesh points to approximate the solution in some near optimal way as the solution changes in time [6] or they add and remove mesh points as needed to minimize the work required to approximate the solution within some prescribed accuracy [7]. This is done by equidistributing a mesh function that reflects the errors in a calculation on a given grid.

BOUNDARY CONDITIONS

Before calculating the solution to any differential equation, boundary conditions should be consistent to form a well-posed problem. A numerical method cannot generate reasonable results for a problem that does not have a well-defined reasonable solution. The importance of proper boundary conditions cannot be overstressed: boundary conditions exert one of the strongest influences on the behavior of the solution. Also, the errors introduced into the calculation from improper boundary conditions persist even as the mesh spacing tends to zero.

A common error in prescribing boundary conditions for hyperbolic equations is to over- or underspecify the number of boundary conditions. Overspecification usually causes nonsmooth solutions with mesh oscillations near the boundary. Underspecification does not ensure a unique solution, and the numerical solution may tend to wander in steady-state calculations. In either case, the results are not accurate and one should be skeptical of even the qualitative behavior of the solution.

It should be noted that the way in which boundary conditions are specified for the difference equations can change a well-posed continuous problem into an ill-posed (unstable) discrete problem. However, our experience with the suggested implementations of following technique has been favorable.

Two of the most common methods used to incorporate boundary conditions into discrete equations are the extrapolation and uncentered difference methods [4].

Extrapolation Method

In this method, the domain of the problem is extended and the solution is extrapolated to fictitious points outside the integration region. The nonphysical solution at these points is defined so that the discrete equations are consistent with as many relationships as can be derived from the boundary conditions and differential equations. The extrapolation formula can do this best by incorporating the discrete boundary conditions [Eq. (3)] into the extrapolant. Additional relations can be generated by differentiating Eq. (2) with respect to time, replacing all time derivatives by space derivatives using Eq. (1), and discretizing the resulting equations.

For example, use a centered five-point formula to approximate the spatial derivative in the equation

$$U_t = (U^2)_{xx}, \quad x \in [0,1], \quad (15)$$

with the boundary conditions

$$B(u(x,t),t) = 0, \quad x = 0,1. \quad (16)$$

The five-point formula at $x_1 = \Delta x$ uses the nonexistent mesh point $x_{-1} = -\Delta x$. The value $U(x_{-1})$ needed here, is constructed explicitly with an extrapolation formula as follows; Eq. (16) is differentiated with respect to time to give

$$\frac{\partial B}{\partial u} (U(0,t),t) U_t + \frac{\partial B}{\partial t} (U(0,t),t) = 0$$

or

$$\frac{\partial B}{\partial u} (U^2)_{xx} + \frac{\partial B}{\partial t} = 0. \quad (17)$$

Discretizing Eq. (17) with Eq. (6) and rearranging yields the extrapolation formula

$$U_{-1} = \text{known data} + O(\Delta x^4). \quad (18)$$

Uncentered Differences

The second approach is to extend the number of boundary conditions so that all components of the solution are defined at the boundary. Again, these additional boundary conditions must be consistent with the original problem and as many relationships as can be derived from it. An uncentered difference approximation is then used to approximate the spatial derivatives at the mesh points nearest the boundary.

This method is described for the linear hyperbolic system of M equations

$$u_t = h(x)u_x, \quad (19)$$

with the boundary conditions

$$su_0 = b(t), \quad x = x_0. \quad (20)$$

Difficulties arise in defining the solution at the boundary when $0 < \text{Rank}(s) < \text{Rank}(h) = M$, and there is no unique solution u_0 of Eq. (20). If $\text{Rank}(s) = 0$, all the characteristics are outgoing, and using either uncentered differences at the points near the boundary or straightforward polynomial extrapolation to the fictitious points gives accurate results. When $\text{Rank}(s) = M$, all the characteristics are entering the boundary and all the solution components can be solved for on the boundary. Uncentered spatial differences can then be used at the points near the boundary and will result in an accurate approximation of the boundary conditions. When $\text{Rank}(s)$ is greater than zero but less than M , then by differentiating Eq. (20) with respect to time and replacing u_t from Eq. (19), we have

$$sh(x)u_x = b'(t), \quad x = x_0. \quad (21)$$

Approximating U_x by second-order one-sided differences gives us

$$SH_0 U_0 = [SH_0(4U_1 - U_2) - 2\Delta x b'(t)]/3 + O(\Delta x^3), \quad (22)$$

where $H_0 = H(x_0)$. Equation (22) gives additional information about the boundary conditions that is consistent with both the original boundary conditions of Eq. (20) and the differential Eq. (19). If we still do not have enough boundary conditions to solve for U_0 uniquely, we can continue by differentiating Eq. (21) with respect to time and using Eq. (19) again.

Often, H_0 is nonlinear and the above procedure must be iterated. Usually one or two iterations will supply a stable accurate boundary approximation.

Once U_0 has been found, we can use uncentered finite differences to approximate the spatial derivatives at the mesh points near the boundary or we can extrapolate the solution to fictitious points outside the integration region by replacing the derivatives in Eq. (21) with second-order centered differences and solving for U_{-1} .

TIME DISCRETIZATION

The numerical solution of Eq. (2) is advanced in time in discrete steps that vary depending on the local behavior of the solution; that is, the length of the time steps depends on whether the solution is evolving on a slow or fast time scale. The methods that approximate the time derivatives, like those that approximate the space derivatives, are based on Taylor series. The major difference between time and space differentiation is that time has a direction. This time flow allows savings in computer storage, but introduces questions about the time stability of the difference equations relative to the stability of the differential equation.

The integration methods discussed in this section are called k -cycle Runge-Kutta multistep methods and can be written

$$U^{n+a_1} = \sum_{i=0}^{k_1} \alpha_{1,i} U^{n-i} + \Delta t \sum_{i=0}^{k_2} \beta_{1,i} F^{n-i} + \Delta t \sum_{i=1}^k \gamma_{1,i} F^{n+a_i}$$

$$\begin{aligned} & \cdot \\ & \cdot \\ & \cdot \end{aligned} \quad (23)$$

$$U^{n+a}_k = \sum_{i=0}^{k_1} \alpha_{k,i} U^{n-i} + \Delta t \sum_{i=0}^{k_2} \beta_{k,i} F^{n-i} + \Delta t \sum_{i=1}^k \gamma_{k,i} F^{n+a}_i$$

$$U^{n+1} = U^{n+a}_k$$

Here, U^{n+a} and $F^{n+a} = F(U^{n+a})$ are approximations to the solution at time $t_{n+a} = t_n + a(t_{n+1} - t_n) = t_n + a\Delta t$, where a is a scalar.

The method is explicit if $\gamma_{i,j} = 0$ for $j \geq i$ and implicit otherwise. Implicit methods require solving one or more algebraic systems on each time step. The extra work required to solve these systems is often rewarded by a substantially larger stability region than a similar explicit method might have.

Explicit Methods

The simplest integration method, called the forward Euler method

$$U^{n+1} = U^n + \Delta t F(U^n) \quad (24)$$

is linearly stable if Δt is chosen so that $\lambda \Delta t$ lies within the stability region shown in Fig. 2. Here, λ is any of the eigenvalues of the linearized Jacobian matrix of F .

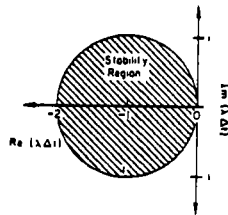


Figure 2
Stability Region for the Forward Euler Method Eq. (24)

If F in Eq. (25) is a second-order finite difference approximation [Eq. (6)] to the parabolic equation

$$u_t = su_{xx} \quad (25)$$

with periodic boundary conditions and an equally spaced mesh between $[-\pi, \pi]$,

$$\lambda = -\frac{2s}{\Delta x^2}(1 - \cos\theta) \quad (26)$$

where θ takes on discrete values between 0 and 2π [3]. From Eq. (26) and

Fig. 2, it follows that the stability restriction for Δt is

$$|\Delta t \lambda_{\max}| \leq 2 \quad , \quad (27)$$

or

$$\frac{\Delta t}{\Delta x} s \leq \frac{1}{2} \quad . \quad (28)$$

This restriction is called the Courant-Friedricks-Lewy or CFL condition. We can vary Δt to retain a stable numerical solution as long as Eq. (28) is not violated.

Equation (24) is a first-order integration method. As in the spatial discretization section, higher order methods are often more cost-effective. If Eq. (24) is used to predict U^{n+1} , then this value can be corrected to second order using the improved Euler corrector method.

$$U^{n+1} = U^n + \frac{\Delta t}{2} (\tilde{F}^{n+1} + F^n) \quad . \quad (29)$$

An excellent explicit method for hyperbolic equations is the second-order leap-frog predictor method

$$\tilde{U}^{n+1} = U^{n-1} + 2\Delta t F^n \quad , \quad (30)$$

combined with the third-order leap-frog correction method [4]

$$U^{n+1} = \frac{4}{5}U^n + \frac{1}{5}U^{n-1} + \frac{2\Delta t}{5}\tilde{F}^{n+1} + \frac{4\Delta t}{5}F^n \quad . \quad (31)$$

The stability region shown in Fig. 3 is very good along the imaginary axis. Note that if one stops after the predictor, the method is unstable when the λ have nonzero real parts, making it unsuitable for parabolic equations.

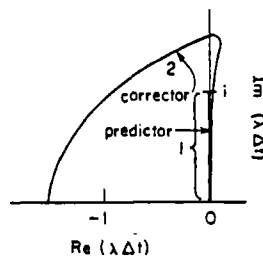


Figure 3
Stability Regions for Leap-frog Predictor Corrector Method

The stability for a centered finite difference approximation to

$$u_t = v u_x \quad (32)$$

with periodic boundary conditions is determined by [3]

$$\lambda_2 = i \sin \theta / \Delta x \quad (33)$$

or

$$\lambda_4 = i(8 \sin \theta - \sin 2\theta) / 6 \Delta x, \quad (34)$$

where λ_2 and λ_4 are the eigenvalues associated with the second- and fourth-order finite difference approximation Eqs. (4) and (5).

The resulting stability restrictions for the second-order space difference are

$$\frac{\Delta t}{\Delta x} v \leq 1 \quad \text{predictor stability [Eq. (30)]} \quad (35)$$

$$\frac{\Delta t}{\Delta x} v \leq \frac{3}{2} \quad \text{predictor-corrector stability [Eqs. (30) and (31)]} \quad (36)$$

For the fourth-order space difference [Eq. (5)] they are

$$\frac{\Delta t}{\Delta x} v \leq \frac{3}{4} \quad \text{predictor stability [Eq. (30)]} \quad (37)$$

$$\frac{\Delta t}{\Delta x} v \leq \frac{9}{8} \quad \text{predictor-corrector stability [Eqs. (30) and (31)]} \quad (38)$$

The leap-frog predictor is unstable for systems of equations with eigenvalues having a nonzero real part. Therefore, when artificial dissipation is added in shock calculations [4], or when the boundary conditions shift the spectrum of the discretized equation, the leap-frog method cannot be used without the corrector cycle. The first corrector application extends the limit on the maximum time step by 50% and improves the method to third order.

Another difficulty of the leap-frog predictor is a unique type of error caused by time and space mesh decoupling. The odd and even points of a mesh are only weakly coupled for even spatial derivatives; errors with frequency $2\Delta x$ or $2\Delta t$ can degrade the accuracy of the solution with high-frequency noise. The corrector cycle couples the mesh points among the three time levels and prevents this instability.

A relatively new explicit method, called the iterated multistep (IMS) method [8], can be written as

$$U^{n+a_i}_i = U^{n+a_{i-1}} + \Delta t c_i [F^{n+a_{i-1}} - F^{n+a_{i-2}}] \quad (39)$$

Here, $a_i = 1$ for $i \geq 3$.

An explicit predictor-corrector is used to start the iteration and the c_i are chosen so that Eq. (39) increases the order of the method on each iteration for linear autonomous systems. For example, if Eqs. (24) and (29) start the iteration, then $c_i = 1/i$, $i = 3, 4, \dots$. If the leap-frog predictor-corrector Eqs. (30) and (31) start the iteration, then $c_3 = 3/10$, $c_4 = 7/30$, $c_5 = 4/21, \dots$.

In general, the stability of the IMS methods increases on every iteration, as can be seen in Fig. 4. Another advantage is that IMS methods allow for local improvements in the stability and accuracy of the calculation. Only a single time level is used in the iteration, so only those ODE components that have failed to pass some accuracy test need be iterated. That is, by iterating locally in regions of rapid changes such as shock fronts, boundary layers, or regions with a refined mesh, the stability and accuracy of the calculation are improved precisely where needed.

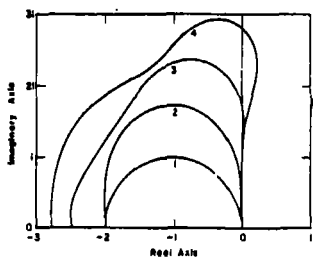


Figure 4a

Stability region for the
Improved Euler IMS

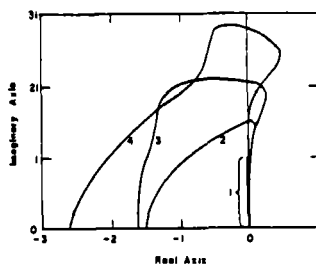


Figure 4b

Stability region for the
Leap-Frog IMS

Implicit Methods

Many problems occur when the solution changes on a slow time scale but the stability criteria limit the time step far below that needed to retain accuracy. In these cases, it is often best to use a more stable implicit method.

Two of the best implicit methods are the second-order trapezoidal rule

$$U^{n+1} - \frac{\Delta t}{2} F^{n+1} = U^n + \frac{\Delta t}{2} F^n \quad (40)$$

and the second-order backward difference (BDF) formula

$$U^{n+1} - \frac{2}{3}\Delta t F^{n+1} = \frac{4}{3}U^n - \frac{1}{3}U^{n-1} \quad (41)$$

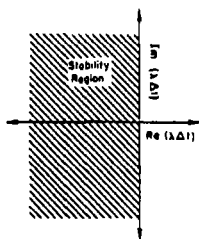
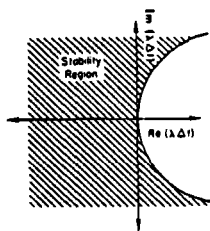
Figure 5a
Trapezoidal MethodFigure 5b
Second-order BDF

Figure 5
Stability Regions for Trapezoidal Method Eq. (40)
and Second-Order BDF Eq. (41) Method

On each time step of a one-cycle implicit method, we must solve a nonlinear algebraic system of the form

$$U^{n+1} + \Delta t \gamma F(U^{n+1}) = \text{known quantities} \quad (42)$$

Several iterative methods, discussed in the next section, show how Eq. (42) might be solved. A good first guess can often be made by using an explicit or extrapolation method.

ALGEBRAIC SYSTEMS

Iterative Methods

In the discussions on space and time discretization, it became necessary to solve large sparse algebraic systems of equations, which can be written

$$A(v) - b = 0 \quad (43)$$

where A is a nonlinear discrete operator, b is a known vector, and the discrete solution vector is v .

Often the solution of Eq. (43) is difficult to obtain directly, but the residual error

$$r = A(w) - b \quad (44)$$

for an approximate solution w is easy to evaluate. If there is a related system

$$P(w) - b = 0 \quad (45)$$

that approximates Eq. (43) and is easier to solve, the defect correction algorithm may be appropriate.

Given a guess v_n near a root v_{n+1} of Eq. (43), we can expand Eq. (43) using Taylor series to get

$$\begin{aligned} 0 &= A(v_{n+1}) - b \\ &= A(v_{n+1}) - b + P(v_{n+1}) - P(v_{n+1}) \\ &= A(v_n) - b + P(v_{n+1}) - P(v_n) - (J_P - J_A)(v_{n+1} - v_n) + O(\varepsilon^2) \quad (46) \end{aligned}$$

where $\varepsilon = v_{n+1} - v_n$. The defect correction iteration is any $O(\varepsilon)$ approximation to Eq. (46). The simplest such iteration is

$$P(v_{n+1}) = P(v_n) - A(v_n) + b \quad (47)$$

This iteration will converge if v_n and J_P , the Jacobian of P , are near enough to v_{n+1} and J_A , respectively. Table 2 lists some of the more common applications of defect corrections.

The iteration (47) can often be speeded up by using a one-step acceleration parameter ω to give

$$P(v_{n+1}) = P(v_n) - \omega_n [A(v_n) - b] \quad (48)$$

These methods include successive overrelaxation, dynamic alternating direction implicit methods, and damped Newton. Often a two-step acceleration method

$$P(v_{n+1}) = b + \omega_n [P(v_n) - \alpha_n A(v_n)] + (1 - \omega_n)P(v_{n-1}) \quad (49)$$

can speed up the convergence even more. These methods include the Tchebyshev [12] and conjugate-gradient [11] methods.

NONLINEAR TROUBLES

When using linear methods to approximate the solution of well-posed linear equations, one can have some confidence in the results, thanks to Lax's equivalence theorem. This theorem states that, for these problems, a stable consistent approximation is necessary and sufficient for the numerical approximation to converge to the true solution as the mesh is refined.

$P(v_{n+1}) =$	Method
$A(v_n) + J_A(v_n)(v_{n+1} - v_n)$	Newton [9]
Diagonal of J_A	Jacobi [9]
Lower triangular part of J_A	Gauss-Siedel [9]
Lower triangular part of J_A + first upper off-diagonal	Line Gauss-Siedel [9]
Coarse grid operator + relax using one of the above	Multigrid [10]
Symmetric part of J_A	Concus-Golub-O'Leary [11]
If $A = (I + \Delta t L_x + \Delta t L_y)$, then $\tilde{P} = (I + \Delta t \tilde{L}_x)(I + \Delta t \tilde{L}_y)$, where \tilde{L}_x = linearized lower order approximation to L .	ADI [9]
LU where L = lower triangular matrix U = upper triangular matrix	Incomplete LU method [9]

Table 2. Common examples of the defect correction iteration.

The theorem is false for general nonlinear equations or nonlinear methods, as shown in the following two examples.

Nonlinear Equation Example

Consider the nonlinear diffusion equation

$$u_t = (u^3)_{xx}, \quad u(x,0) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases} \quad (50)$$

The solution to this equation is a wave front traveling to the right.

Equation (50) can be rewritten as

$$u_t = 3u(u u_{xx} + 2u_x^2) \quad (51)$$

The solution U is discretized and the space derivatives are approximated by any method in the Space Discretization section. The time variable is integrated with any stable method from the Time Discretization section.

Note that $U_t(x_i) = 0$, a discrete version of Eq. (51), is zero for all t when $x_i > 0$ since $U_i = 0$ at these mesh points. This implies that the wave can never propagate to the right of zero, no matter how fine a mesh is used.

Thus, we have a stable consistent approximation that can never converge to the true solution. If Eq. (50) is differenced directly using Eq. (6) with $G = U^3$ then the resulting numerical solution will converge to the correct answer.

Nonlinear Method Example

Consider the unidirectional wave equation

$$u_t = u_x, \quad u(x,0) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases} \quad (52)$$

with the solution $u(x,t) = u(x+t,0)$.

After discretizing, we approximate the space derivative using a nonlinear transformation method (Space Discretization, Grid Resolution). The transform interpolates the discrete solution with a monotonicity preserving interpolant that has a continuous first derivative [13]. The derivative of this interpolant is then used at the grid points to define the ODEs [Eq. (3)].

We know that the true solution to Eq. (52) is monotone and we might expect that an interpolant that preserves this monotonicity will greatly improve the accuracy of the calculation. However, since the interpolant is monotone, $U(x_i) = 0$, we have $U_t(x_i) = 0$ at all the grid points. The solution never changes!

SUMMARY

We have used a modular approach to develop accurate and robust methods for the numerical solution of PDEs. The methods to discretize the spatial operator, the boundary conditions, and the time variable, and solve any algebraic system that may arise are combined when writing a code to solve the PDE system. As the last two examples show, special care must be taken when solving a nonlinear equation or when using a nonlinear method. This means that the code must be field tested.

The field test is to check the reliability of the method on a particular nonlinear system of PDEs. The numerical results should be insensitive to reformulations of the equations, small changes in the initial conditions, the mesh orientation and refinement, and the choice of a stable accurate discretization method.

Another excellent analysis tool is verification that any auxiliary relationships (such as conservation laws) hold for the numerically generated solution. These checks should be made - even if one is absolutely, positively sure that the numerical solution and coding are correct.

When the above checks are made, the methods given here can be combined to give reliable efficient methods for solving a fairly large class of nonlinear PDEs. Also, by using the modular approach presented here, these codes can evolve efficiently since new methods can be quickly incorporated into the program.

ACKNOWLEDGEMENT

I am grateful to Joe Dendy, Blair Swartz, and Burton Wendroff for providing much welcome advice in our many discussions during this work.

REFERENCES

- [1] Wachspress, E. L., Iterative solution of elliptic systems (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1966).
- [2] Ciment, M., Leventhal, S., and Weinberg, B., "The operator compact implicit method for parabolic equations," J. Comp. Phys. 28 (1978) 135-166.
- [3] Hyman, J. M., The method of lines solution of partial differential equations, Courant Institute of Mathematical Sciences report C00-3077-139 (1976).
- [4] Hyman, J. M., A method of lines approach to the numerical solution of conservation laws, Adv. in Comp. Methods for PDEs - III, Vichnevetsky, R., and Stepleman, R. S., (eds) Publ. IMACS (1979) 313-321.
- [5] Hyman, J. M., "The numerical solution of time dependent PDEs on an adaptive mesh," Los Alamos Scientific Laboratory LA-UR-80-3702 (1980).
- [6] Gelinas, R. J., Doss, S. K., and Miller, K., "The moving finite element method: applications to general partial differential equations with multiple large gradients," J. Comp. Phys. 40 (1981) 202-249.
- [7] Berger, M., Gropp, W., and Oliger, J., Grid generation for time dependent problems: criteria and methods, numerical grid generation techniques, NASA Conf. Proc. Publication 2166 (1980) 181-188.
- [8] Hyman, J. M., Explicit A-stable methods for the solution of differential equations, Los Alamos Scientific Laboratory LA-UR-79-29 (1979).
- [9] Young, D. M., Iterative solution of large linear systems (Academic Press, New York, NY, 1971).
- [10] Brandt, A., Multi-level adaptive solutions to boundary value problems, Math. Comp. 31 (1977) 333-390.

- [11] Concus, P., Golub, G. H., and O'Leary, D. P., Numerical solution of nonlinear elliptic partial differential equations by the generalized conjugate gradient method, *Computing* 19 (1978) 321-339.
- [12] Manteuffel, T. A., The Tchebychev iteration for nonsymmetric linear systems, *Numer. Math.* 28 (1977) 307-327.
- [13] Hyman, J. M., Accurate monotonicity preserving cubic interpolation, Los Alamos National Laboratory report LA-8796-MS (1981).

This Page Intentionally Left Blank

LIMIT ANALYSIS OF PHYSICAL MODELS

M. FREMOND

Laboratoire Central des Ponts et Chaussées
58 bld Lefebvre - 75732 PARIS Cedex 15 - France -

A. FRIAÂ

Ecole Nationale d'Ingénieurs de Tunis
B.P. n°37, Belvédère - TUNIS - Tunisie -

M. RAUPP

Centro Brasileiro de Pesquisas Fisicas
avenida Wenceslau-Braz 71, CEP 2000, RIO DE JANEIRO - Brésil -

Let us consider a physical model where some internal variables must belong to a given set in order for the model to be valid or not to breakdown. Our aim is to determine the external actions which achieve this condition.

We present two examples. The first one describes a simple case in limit analysis of structures in mechanics. The second describes the breakdown analysis of an electrical circuit.

In conclusion a common formalism is presented which can be applied to other situations [2].

AN EXAMPLE IN MECHANICS [2]

Let us consider the beam described on figure 1.

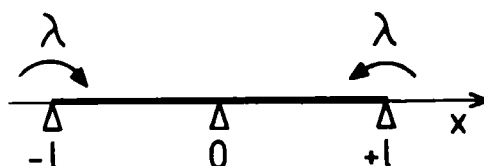


Figure 1

The beam loaded by the torque λf .

Two equal and opposite torques, with intensity λ , are applied at its ends.

The interior forces of the beam are the bending moment at any point x of the beam. We assume the beam is made of a material which cannot support moments greater than k . This threshold k can have several physical meanings, elastic limit, yielding, ...

We look for the external loads, characterized by the number λ , such that there exist a moment which equilibrate them and which are lower than the threshold k .

Let us recall several notions, most of them being classical in the strength of material theory.

The set of kinematically admissible displacements $v(x)$ is

$$V = \{v \mid v \in W^{2,1}(-\ell, +\ell) ; v(-\ell) = v(0) = v(\ell) = 0\}$$

with the norm $\|v\| = \int_{-\ell}^{+\ell} |v''(x)| dx$.

We use throughout this text the Sobolev spaces $W^{2,1}$, H^1 , H^2 , ... which describe the regularity of the functions involved in the equations [1].

The set of the external loads is V' the dual space of V . The applied torques λf are, indeed, elements of V' because

$$\langle \lambda f, v \rangle = \lambda (v'(\ell) - v'(-\ell)) = \lambda \int_{-\ell}^{+\ell} v''(x) dx,$$

(the double brackets denote the duality application between V and V').

The set of the strain rates D and the set of the interior forces S are respectively

$$D = L^1(-\ell, +\ell), \quad S = D' = L^\infty(-\ell, +\ell).$$

The deformation operator ε , which is linear and continuous from V into D , is defined by

$$\varepsilon(v) = v''.$$

Let us call ${}^t\varepsilon$ the transposed operator of ε .

The mechanical power developed by the moment M and the strain rate d is

$$- \langle M, d \rangle = - \int_{-\ell}^{+\ell} M(x) d(x) dx.$$

The set $\tilde{\Sigma}$ of statically admissible bending moments, i.e. the set of M which equilibrate an exterior load λf , is

$$\begin{aligned} \tilde{\Sigma} &= \{M \in S \mid \exists \lambda \in \mathbb{R} ; \forall v \in V, -\langle M, \varepsilon(v) \rangle + \langle \lambda f, v \rangle = 0 ; \text{ or } {}^t\varepsilon(v) = \lambda f\} \\ &= \{M \in S \mid M'' = 0 \text{ on }]-\ell, 0[\cup]0, \ell[, M \text{ is continuous at } 0 ; M(-\ell) = M(\ell) = \lambda\}. \end{aligned}$$

Let us remark that $\forall M \in S, \forall v \in V, \langle M, \varepsilon(v) \rangle = \langle \lambda f, v \rangle = \lambda (v'(\ell) - v'(-\ell))$.

This relation defines two linear applications $\tilde{L} \in \mathcal{L}(\tilde{\Sigma}, \mathbb{R})$, $L \in \mathcal{L}(V, \mathbb{R})$ by

$$\tilde{L}(M) = \lambda \text{ and } L(v) = v'(\ell) - v'(-\ell).$$

The set G of the moments which can be supported is

$$G = \{M \mid M \in S ; |M(x)| \leq k \text{ a.e.}\}.$$

The problem we want to solve is ;

$$\text{Find } K \subset \mathbb{R} \text{ such that } \lambda \in K \iff \exists M \in G \cap \tilde{\Sigma} \text{ such that } \tilde{L}(M) = \lambda.$$

It turns out that

$$K = \tilde{L}(G \cap \tilde{\Sigma})$$

and that under reasonable assumptions K is a closed convex set [5]. This set K describes the loads which have the possibility to be supported by the structure or for which the structure can be safe. The loads which are not in K cannot be supported. Within the framework of plasticity the boundary of K is the set of the limit loads [7].

The convex set K is of fundamental interest in practice. There are two ways to compute it : the static and kinematic methods.

The static method.

It uses the fact that the set K is convex. If it is possible to find $M \in G \cap \tilde{\Sigma}$, the line $[0, L(M)]$ is contained in K (figure 3). Here this method leads to $K = [-k, +k]$. Unfortunately in many practical applications it is impossible to find elements of $G \cap \tilde{\Sigma}$ (for instance in 2 or 3 dimensions problems) and this method fails.

The kinematic method.

It uses convex analysis results. Let us consider the support function Π_G of the convex set G :

$$\Pi_G(d) = \text{Sup}\{ \langle M, d \rangle \mid M \in G \} = k \int_{-\ell}^{+\ell} |d(x)| dx,$$

and the function

$$\begin{aligned} Q \in \mathbb{R} \longrightarrow \psi_{K_1}(Q) &= \text{Sup}\{ Q \cdot L(v) - \Pi_G(\varepsilon(v)) \mid v \in V \} = \\ &= \text{Sup}\{ Q \int_{-\ell}^{+\ell} v''(x) dx - k \int_{-\ell}^{+\ell} |v''(x)| dx \mid v \in V \}. \end{aligned}$$

It is easy to show that $\psi_{K_1}(Q)$ is either equal to $+\infty$ if $Q > k$ or to 0 if $Q \leq k$. In general ψ_{K_1} is the indicator function of a convex set K_1 . Under reasonable assumptions K equals K_1 [2], [3], [5].

Let us notice that

$$K_1 \subset \{ Q \mid Q \cdot L(v) - \Pi_G(\varepsilon(v)) \leq 0 \text{ for any } v. \}$$

This is the key point of the kinematic method (figure 4) and we have

$$\begin{aligned} K_1 &= \{ Q \mid Q \leq \text{Inf}\{ \Pi_G(\varepsilon(v)) \mid v \in V ; L(v) = 1 \} \cap \dots \\ &\quad \{ Q \mid Q \leq \text{Inf}\{ \Pi_G(\varepsilon(v)) \mid v \in V ; L(v) = -1 \} \}. \end{aligned}$$

The practical steps of the method are the computation of Π_G and minimisation problems. In mechanics this kinematic method is often more efficient.

AN EXAMPLE IN ELECTRICITY [2] [4].

Let us consider the electrical circuit described on the figure 2. A cell whose constitutive law is shown on figure 2 is connected in parallel with the condenser. This cell blows if the potential difference v_2 between its two connections exceeds the given threshold v_c . At the time $t = 0$, the circuit being at rest (the condenser is discharged and there is no current through it) and electromotive force

$$e(t) = \sum_{n=1}^N \lambda_n e_n(t)$$

is applied. The problem is to find the controls $\lambda = (\lambda_1, \dots, \lambda_N)$ such that the cell does not blow during the time interval $[0, T]$. This problem is related to the preceding one but is no longer static.

Let us define, v_1 the potential difference between the connections of the resistance and inductance block, j_1 the current intensity inside this block, j_2 the

current intensity sum of the current intensities a_2 and b_2 which pass through the condenser and the cell, i the current intensity which passes through the electromotive force.

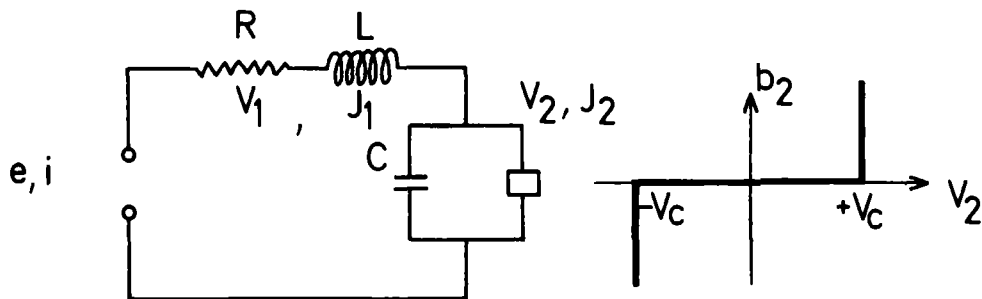


Figure 2

The circuit and the constitutive law of the cell.

We have the relations [6]

$$\sum_{n=1}^N \lambda_n e_n = v_1 + v_2, \quad (1)$$

$$v_1 = Rj_1 + L \frac{dj_1}{dt}, \quad (2)$$

$$a_2 = C \frac{dv_2}{dt}, \quad (3)$$

$$b_2 \in \partial\psi(v_2), \quad (4)$$

where $\partial\psi$ is the subdifferential of the indicator function of $[-v_c, +v_c]$,

$$i = j_1 = j_2 = a_2 + b_2, \quad (5)$$

$$v_2(0) = 0, \quad \frac{dv_2}{dt}(0) = 0, \quad (6)$$

Our problem is to find the set $K \subset \mathbb{R}^N$ of the controls λ which can be supported by the circuit : i.e. for which the cell does not blow,

$$\lambda \in K, \iff \forall t \in [0, T], \quad b_2(t) \neq 0.$$

Let us define some sets which are useful to describe K :

$I = H^2(0, T)$, the set of the current intensities inside the electromotive force,

$E = H^1(0, T)$, the set of the electromotive forces,

the electrical work developed by e and i is

$$\langle e, i \rangle = \int_0^T e(t) i(t) dt,$$

$J_1 = J_2 = H^2(0, T)$, the sets of current intensities j_1 and j_2 ,

$V_1 = H^1(0, T)$, $V_2 = \{v_2 \in H^3(0, T); v_2(0) = \frac{dv_2}{dt}(0) = 0\}$, the sets of the difference of potential v_1 and v_2 ,

$$J = J_1 \times J_2, \quad V = V_1 \times V_2,$$

the electrical work developed by $j = (j_1, j_2)$ and $v = (v_1, v_2)$ is

$$\langle v, j \rangle = \int_0^T (j_1 v_1 + j_2 v_2) dt.$$

The first Kirchoff law (equilibrium of the nodes) is

$$(j_1, j_2) = (i, i) \text{ or } \varepsilon(i) = j, \text{ with } \varepsilon(i) = (i, i).$$

The operator ε is linear and continuous from I into J .

The second Kirchoff law (equilibrium of the loops) is

$$\forall i \in I, \langle v, \varepsilon(i) \rangle = \langle e, i \rangle \text{ or } e = i_1 + i_2 \text{ or } {}^t \varepsilon(v) = e,$$

where ${}^t \varepsilon$ is the transposed operator of ε .

We define the set \tilde{J} of the potential differences v which equilibrate a control λ

$$\tilde{J} = \{v \in V \mid \exists \lambda \in \mathbb{R}^N; {}^t \varepsilon(v) = v_1 + v_2 = \sum_{n=1}^N \lambda_n e_n\}$$

and the set G of the potential differences which can be supported by the elements of the circuit :

$$G = \{v \in V \mid \exists i \in I; \varepsilon(i) = j; v_1 = R j_1 + L \frac{dj_1}{dt}; j_2 = C \frac{dv_2}{dt}; |v_2| \leq v_c\}.$$

One can remark that

$$\forall i \in I, \forall v \in \tilde{J}, \langle v, \varepsilon(i) \rangle = \langle e, i \rangle = \sum_{n=1}^N \lambda_n \langle e_n, i \rangle$$

This relation defines two linear applications $\tilde{L} \in \mathcal{L}(\tilde{J}, \mathbb{R}^N)$ and $L \in \mathcal{L}(I, \mathbb{R}^N)$ by

$$\tilde{L}(v) = \lambda \text{ and } L(i) = (\langle e_n, i \rangle).$$

It turns out that

$$K = \tilde{L}(G \cap \tilde{J}).$$

The formulation of both electrical and mechanical problems are exactly the same. It is then possible to use the same static and kinematic methods to find K .

The "static" method

Let us choose $\lambda \in \mathbb{R}^N$ and solve the set of equations (1,2,3,5,6) with $b_2 = 0$. If the solution is such that $|v_2(t)| \leq v_c$ for any t , the line $[0, \lambda]$ is in K , if not the line $[\lambda, +\infty]$ is not in K (figure 3).^c

The "kinematic" method

Let us define the support function Π_G of G and compute

$$\Pi_G(\varepsilon(i)) = v_c \int_0^T \left| i - CR \frac{di}{dt} + \frac{d^2 i}{dt^2} \right| dt + v_c LC \left| \frac{di}{dt}(T) \right| \text{ if } i(T) = 0,$$

$$\Pi_G(\varepsilon(i)) = +\infty, \text{ if } i(T) \neq 0.$$

The convex set K_1 is defined by its indicator function

$$\psi_{K_1}(\lambda) = \text{Sup}\{\lambda \cdot L(i) - \Pi_G(\varepsilon(i)) \mid i \in I\}.$$

Under reasonable assumptions [3], we have $K = K_1$. The key point of the method is then

$$K = K_1 \subset \{\lambda \mid \lambda \cdot L(i) - \Pi_G(\varepsilon(i))\} \text{ for any } i.$$

Let us choose a direction q of \mathbb{R}^N , we have

$$K \subset \{\lambda \mid \lambda \cdot q - \text{Inf}\{\Pi_G(\varepsilon(i)) \mid i \in V; L(i) = q\}\}.$$

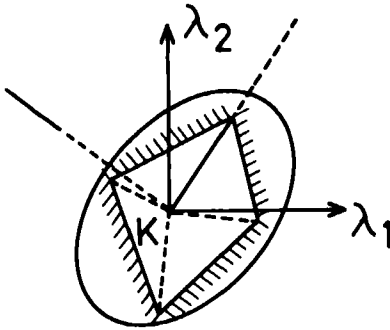


Figure 3

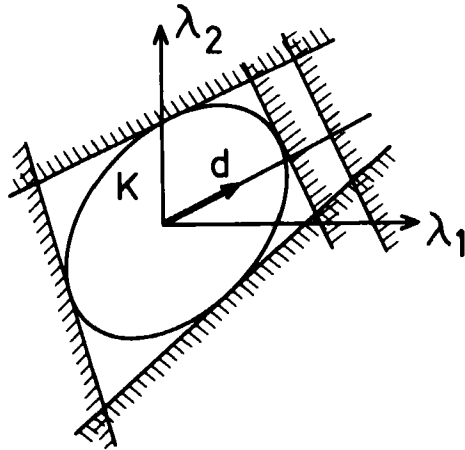


Figure 4

The static method approaches K by the interior and the kinematic method by the exterior.

The two step of this methods are again the computation of Π_G and minimisation problems.

An example

Let us assume that $R=0$, $N=1$, $\forall t$, $e(t)=1$, and define $\omega = 1/\sqrt{LC}$. It is possible to show by the kinematic method (i.e. without solving a differential equation) that $K = [-\lambda_0, +\lambda_0]$ with $\lambda_0 = \frac{v_c}{2}$ if $\omega T \geq \pi$ and $\frac{v_c}{1 - \cos \omega T}$ if $\omega T \leq \pi$.

It is natural that the limit electromotive force to be big for small T because the circuit is at rest at the time $t=0$ and the potential difference v_2 is continuous.

FORMALISM [2], [5], [7].

The electrical and mechanical problem are both illustration of a general formalism based on the classical framework of physics and continuum mechanics.

Let us consider a structure which is described by interior variables $d \in D$ and

$s \in S = D'$ (dual space of D) which develop the power (or work) $\langle d, s \rangle$ and by exterior variables $v \in V$ and $f \in F = V'$ (dual space of V) which develop the power (or work) is $\langle f, v \rangle$. Let us also consider the linear and continuous operator ε from V into D whose transposed operator t_ε is defined by

$$\forall v \in V, \forall s \in S, \langle \varepsilon(v), s \rangle = \langle v, t_\varepsilon(s) \rangle.$$

We say that the interior variable s equilibrates f if $t_\varepsilon(s) = f$.

We assume that the structure can support only the interior variables which are inside a convex set $G \subset S$ and that it is submitted to a loading process [7] defined by a finite dimension subspace F_n of F ($f \in F_n \iff f = \sum_{i=1}^n \lambda_i f_i$; $(\lambda_i) = \lambda \in \mathbb{R}^n$)

The practical problem is to find the set K of \mathbb{R}^n such that for any $\lambda \in K$, the external load $\sum_{i=1}^n \lambda_i f_i$ is equilibrated by an internal variable s which is in G .

We define the set of the internal variables which equilibrate an external load of F_n by

$$\tilde{\Sigma} = \{s \in S \mid \exists \lambda \in \mathbb{R}^n; t_\varepsilon(s) = \sum_{i=1}^n \lambda_i f_i\}.$$

We have

$$\forall s \in \tilde{\Sigma}, \forall v \in V, \langle \varepsilon(v), s \rangle = \sum_{i=1}^n \lambda_i \langle f_i, v \rangle.$$

This relation defines two applications \tilde{L} and L by

$$\tilde{L}(s) = \lambda \text{ and } L(v) = (\langle f_i, v \rangle).$$

It turns out that

$$K = \tilde{L}(G \cap \tilde{\Sigma})$$

and that under reasonable assumptions both static and kinematic methods can be used to compute K [3]. The static method requires the knowledge of elements of $G \cap \tilde{\Sigma}$. The kinematic method requires the computation of the support function Π_G of G and the minimisation of a function.

Let us also remark that the static method is an approach of K by the interior the kinematic method is an approach by the exterior (figures 3 and 4).

REFERENCES

- [1] Adams, R., Sobolev Spaces (Academic Press, London, 1975).
- [2] Frémond, M., Méthodes variationnelles en calcul des structures (Cours de D.E.A. Ecole Nationale des Ponts et Chaussées, Paris, 1979).
- [3] Frémond, M., Friaâ, A., Comptes Rendus de l'Académie des Sciences, 286, Série A, 1978, p. 107.
- [4] Frémond, M., Friaâ, A., Raupp, M., Comptes Rendus de l'Académie des Sciences, 292, Série II, 1981, p. 21.
- [5] Friaâ, A., Le matériau de Norton-Hoff généralisé en élasticité et viscoplasticité. Thèse. Université Pierre et Marie Curie, Paris, 1979
- [6] Raupp, M., Baiocchi, D., Bol. Soc. Mat. Brasil (to appear).
- [7] Salençon, J., Calcul à la rupture et plasticité (Cours de D.E.A. Ecole Nationale des Ponts et Chaussées, Paris, 1978).

This Page Intentionally Left Blank

VISCOSITY SOLUTIONS OF HAMILTON-JACOBI EQUATIONS

Michael G. Crandall*

Department of Mathematics and
Mathematics Research Center
University of Wisconsin-Madison
Madison, Wisconsin

INTRODUCTION

This paper concerns problems of the forms

$$\begin{aligned} H(x, u, Du) &= 0 \quad \text{for } x \in \Omega, \\ u(x) &= z(x) \quad \text{for } x \in \partial\Omega, \end{aligned} \tag{DP}$$

and

$$\begin{aligned} u_t + H(x, t, u, Du) &= 0 \quad \text{for } t > 0, x \in \Omega, \\ u(x, t) &= z(x, t) \quad \text{for } t > 0, x \in \partial\Omega, \\ u(x, 0) &= u_0(x) \quad \text{for } x \in \Omega; \end{aligned} \tag{CP}$$

which we refer to as, respectively, the Dirichlet and Cauchy problems for Hamilton-Jacobi equations. In (DP) and (CP), Ω is an open subset of \mathbb{R}^N , $\partial\Omega$ is the boundary of Ω , the unknown u and the Hamiltonian H are real-valued functions of the relevant arguments and $Du = (u_{x_1}, \dots, u_{x_N})$ indicates the spatial gradient of u . Equations of this sort arise in many contexts, including mechanics, control theory, calculus of variations, differential games, and the approximation of solutions of linear hyperbolic equations.

Analysis of (DP) or (CP) by the classical method of characteristics (see below) is limited to local considerations owing to the fact that characteristics "cross". This phenomena leads to the development of "shocks" - which here roughly means that derivatives develop discontinuities - and then the analysis by characteristics is no longer valid. A global theory of (DP) or (CP) thus requires a notion of solution broad enough to encompass discontinuities in derivatives. To set the stage for the notion introduced herein, we first recall the classical method in a very simple case and make some remarks about the inadequacy of the most obvious way to define generalized solutions of (DP), (CP).

Let us give the classical solution of (CP) in the case where $\Omega = \mathbb{R}$ and $H = H(u_x)$ depends only on u_x . Consider a (sufficiently smooth) function $u(x, t)$ which satisfies $u_t + H(u_x) = 0$ for $0 < t < T$, $x \in \mathbb{R}$. Define $X(s, x_0)$ for $0 < s < T$, $x_0 \in \mathbb{R}$ by

*Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

$$\begin{aligned} \frac{d}{ds} X(s, x_0) &= H'(u_x(X(s, x_0), s)) , \\ X(0, x_0) &= x_0 . \end{aligned} \quad (0.1)$$

We will prove shortly that

$$u_x(X(s, x_0), s) \equiv u_x(X(0, x_0), 0) = u_x(x_0, 0) . \quad (0.2)$$

Then (0.1) implies

$$X(s, x_0) = x_0 + sH'(p_0) \quad (0.3)$$

where $p_0 = u_x(x_0, 0)$. Moreover

$$\frac{d}{ds} u(X(s, x_0), s) = u_x X' + u_t \equiv p_0 H'(p_0) - H(p_0) , \quad (0.4)$$

and we conclude from (0.2)-(0.4) that

$$\begin{aligned} (i) \quad & u_x(x_0 + sH'(p_0), s) \equiv p_0 , \\ (ii) \quad & u(x_0 + sH'(p_0), s) = u(x_0, 0) + s(p_0 H'(p_0) - H(p_0)) . \end{aligned} \quad (0.5)$$

These formulae determine u and u_x from the initial data $u(x, 0)$ so long as the transformation $(x, t) \rightarrow (x + tH'(u_x(x, 0)), t)$ is well-behaved. However, (0.5) (i) says u_x is constant on the lines $x = x_0 + tH'(u_x(x_0, 0))$ and if two of these cross (as they must unless $H'(u_x(x_0, 0))$ is monotone in x_0), we conclude that u_x must develop a discontinuity at the first such crossing. Of course, (0.5) (ii) also yields contradictory information about the values of u when characteristics cross, but discontinuities develop in u_x first. (Roughly, until characteristics cross u_x remains bounded and u therefore cannot develop "jumps".) It remains to verify (0.2). This follows from

$$\frac{d}{ds} u_x(X(s, x_0), s) = u_{xx} X' + u_{xt} = u_{xx} H' + u_{xt} = (u_t + H(u_x))_x = 0 .$$

We now have recalled that in order to have a global time theory for

$$\begin{aligned} u_t + H(u_x) &= 0, \quad t > 0, x \in \mathbb{R} , \\ u(x, 0) &= u_0(x), \quad x \in \mathbb{R} , \end{aligned} \quad (0.6)$$

which extends the classical theory, we must allow discontinuities in u_t and u_x . The simplest way to do this is just to require that u be (locally)

Lipschitz continuous (so u is differentiable at almost every (x, t)) and then ask the equation be satisfied almost everywhere. This simple approach is quite inadequate, for one will not then have uniqueness of solutions of (0.6). Indeed, consider the problem

$$\begin{aligned} u_t - (u_x)^2 &= 0, \quad t > 0, x \in \mathbb{R} , \\ u(x, 0) &= 0, \quad x \in \mathbb{R} . \end{aligned} \quad (0.7)$$

Clearly, $u \equiv 0$ is a solution of (0.7). Another solution, in the almost everywhere sense, is $u = 0$ for $|x| > t$ and $u = t - |x|$ for $t > |x|$. How can one distinguish between these solutions?

Let us motivate the answer to this question given herein. To "solve" (0.6) one frequently uses the approximation of it by the regularized problem

$$\begin{aligned} v_t + H(v_x) - \varepsilon v_{xx} &= 0, \quad t > 0, x \in \mathbb{R}, \\ v(x, 0) &= u_0(x), \quad x \in \mathbb{R}, \end{aligned} \quad (0.6)_\varepsilon$$

in which $\varepsilon > 0$ and considers the limit of the solution of (0.6) as, $\varepsilon \rightarrow 0$. This is an example of the method of "vanishing viscosity". Assume

$v_t + H(v_x) - \varepsilon v_{xx} = 0$. Let $\varphi \in \mathcal{D}(\mathbb{R} \times (0, \infty))^+$ (the nonnegative C_0^∞ functions on $\mathbb{R} \times (0, \infty)$). We have, on the support of φ ,

$$0 = \frac{1}{\varphi} (\varphi v_t) + H\left(\frac{1}{\varphi} \varphi v_x\right) - \varepsilon \frac{1}{\varphi} \varphi v_{xx},$$

and

$$\begin{aligned} \varphi v_t &= (\varphi v)_t - \varphi_t v, \\ \varphi v_x &= (\varphi v)_x - \varphi_x v, \\ \varphi v_{xx} &= (\varphi v)_{xx} - 2\varphi_x v_x - \varphi_{xx} v, \\ &= (\varphi v)_{xx} - \frac{2\varphi_x}{\varphi} (\varphi v)_x + \frac{2\varphi_x^2}{\varphi^2} v - \varphi_{xx} v \end{aligned}$$

so that

$$0 = \frac{(\varphi v)_t}{\varphi} - \frac{\varphi_t}{\varphi} v + H\left(\frac{(\varphi v)_x}{\varphi} - \frac{\varphi_x}{\varphi} v\right) - \varepsilon \left(\frac{(\varphi v)_{xx}}{\varphi} - \frac{2\varphi_x}{\varphi^2} (\varphi v)_x + \frac{2\varphi_x^2}{\varphi^2} v - \frac{\varphi_{xx}}{\varphi} v \right).$$

Now assume $t_0 > 0$, $x_0 \in \mathbb{R}$ and

$$\varphi(x_0, t_0) v(x_0, t_0) = \max_{\substack{x \in \mathbb{R} \\ 0 < t < t_0}} (\varphi(x, t) v(x, t)).$$

Then $(\varphi v)_t(x_0, t_0) > 0$, $(\varphi v)_x(x_0, t_0) = 0$ and $(\varphi v)_{xx}(x_0, t_0) < 0$, so we can evaluate the above identity at (x_0, t_0) to conclude

$$-\frac{\varphi_t}{\varphi} v + H\left(-\frac{\varphi_x}{\varphi} v\right) - \varepsilon \left(\frac{\varphi_{xx}}{\varphi^2} v - \frac{\varphi_{xx}}{\varphi} v \right) < 0$$

at (x_0, t_0) . Assuming that the solution v of $(0.6)_\varepsilon$ tends to a limit u as $\varepsilon \rightarrow 0$ we therefore expect that

$$-\frac{\varphi_t}{\varphi} u + H\left(-\frac{\varphi_x}{\varphi} u\right) < 0 \quad (0.8)$$

on the set where φu assumes its maximum value.

In the next section the notion of solution suggested by these considerations is presented in generality together with a few basic results. Subsequently, in Section 2, the related existence and uniqueness theory is illustrated in model problems. Details and proofs are not given in this brief introductory exposition. They may be found in the paper [1] by the author and P. L. Lions, which is the origin of this theory. We also rely on [1] and P. L. Lions [6] for their references to the very substantial previous literature on the subject of Hamilton-Jacobi equations. Here we only mention that much of this literature is concerned with convex Hamiltonians and our notion of solution in the general nonconvex case appears completely new. There is some parallel, however, with the theory of a single conservation law, e.g. [5], [7]. We should also mention the works of Fleming [3] and Friedman [4] in the nonconvex case, wherein it is shown that the vanishing viscosity method converges. When this method converges, one

has singled out a particular solution and this can be adopted as a (nonintrinsic) definition of a solution. Finally we mention [2] in which error estimates are given which establish the convergence of the vanishing viscosity method as well as a class of difference approximations to our viscosity solutions.

SECTION 1. VISCOSITY SOLUTIONS

Let O be an open subset of \mathbb{R}^M and $F : O \times \mathbb{R} \times \mathbb{R}^M \rightarrow \mathbb{R}$ be continuous. We denote points of O by y and $Du = (u_{y_1}, \dots, u_{y_M})$. The equation

$$F(y, u, Du) = 0 \quad (1.1)$$

includes the equations $H(x, u, Du) = 0$ and $u_t + H(x, t, u, Du) = 0$ by taking $y = x$ in the first case and $y = (x, t)$ in the second. $C(O)$ and $C(O \times \mathbb{R} \times \mathbb{R}^M)$ denote the continuous real-valued functions on O and on $O \times \mathbb{R} \times \mathbb{R}^M$ respectively and convergence in these spaces means uniform convergence on compact sets. $\mathcal{D}(O)^+$ denotes the nonnegative C^∞ functions on O which are supported on compact subsets of O .

If $\psi \in C(O)$ we set

$$E_+(\psi) = \{y \in O : \psi(y) = \max_O \psi > 0\}$$

and

$$E_-(\psi) = \{y \in O : \psi(y) = \min_O \psi < 0\}.$$

If ψ does not assume a positive maximum value on O , then $E_+(\psi) = \emptyset$, etc.

Definition 1. A function $u \in C(O)$ which satisfies

$$\left\{ \begin{array}{l} \text{If } \varphi \in \mathcal{D}(O)^+, k \in \mathbb{R} \text{ and } E_+(\varphi(u - k)) \neq \emptyset, \\ \text{then there is a } y \in E_+(\varphi(u - k)) \text{ for which} \\ F(y, u(y), -\frac{(u(y) - k)}{\varphi(y)} D\varphi(y)) < 0. \end{array} \right. \quad (1.2)$$

is a viscosity subsolution of $F = 0$. Similarly, when u satisfies

$$\left\{ \begin{array}{l} \text{If } \varphi \in \mathcal{D}(O)^+, k \in \mathbb{R} \text{ and } E_-(\varphi(u - k)) \neq \emptyset, \\ \text{then there is a } y \in E_-(\varphi(u - k)) \text{ for which} \\ F(y, u(y), -\frac{(u(y) - k)}{\varphi(y)} D\varphi(y)) > 0. \end{array} \right. \quad (1.3)$$

it is a viscosity supersolution of $F = 0$. If u satisfies both (1.2) and (1.3), it is a viscosity solution of $F = 0$.

We sometimes use "viscosity solution of $F < 0$ " in place of "viscosity subsolution of $F = 0$ ", etc. The line of argument which led from (0.6) to (0.7) (when applied to $u - k$) also leads from sub- and supersolutions of (0.6) to the corresponding "viscosity" notions. We caution the reader here that the viscosity solutions of $F = 0$ are not, in general, the viscosity solutions of $-F = 0$. However, u is a viscosity solution of $F < 0$ if and only if $v = -u$ is a viscosity solution of $-F(x, v, -Dv) > 0$. If $u \in C^1(O)$, then

$$D(\varphi(u - k)) = 0 \quad \text{on} \quad E_+(\varphi(u - k)) \quad \text{so}$$

$$- \frac{(u - k)}{\varphi} D\varphi = Du \quad \text{on} \quad E_+(\varphi(u - k)),$$

etc. It follows that classical solutions of $F < 0$, $F > 0$ and $F = 0$ are also viscosity solutions. It is also true that smooth viscosity solutions are classical solutions, as we see later.

To set the stage for the next result, consider the following problem:

$$\begin{aligned} (u_x)^2 - 1 &= 0, & -1 < x < 1 \\ u(-1) &= u(1) = 0. \end{aligned} \tag{1.4}$$

This problem has a largest Lipschitz continuous solution $u = 1 - |x|$ and a smallest Lipschitz continuous solution $u = |x| - 1$ which satisfy the equation a.e. It has many other solutions, e.g.

$$\begin{aligned} u_n(-1) &= 0, \quad u_n'(x) = (-1)^j \quad \text{on} \quad [(j - 2n)/2n, (j + 1 - 2n)/2n], \\ j &= 0, \dots, 4n - 1, \end{aligned}$$

defines a solution u_n such that $0 < u_n < 1/2n$ for $n = 1, 2, \dots$. Clearly $u_n \rightarrow 0$ uniformly, but $u \equiv 0$ is not a solution of $(u_x)^2 = 1$ anywhere. In contrast, for viscosity solutions we have:

Theorem 1. (Stability of the class of viscosity solutions). Let $\{F_\ell\}$ be a sequence in $C(0 \times \mathbb{R} \times \mathbb{R}^M)$ and u_ℓ be a viscosity solution of $F_\ell = 0$. Let $F_\ell \rightarrow F$ in $C(0 \times \mathbb{R} \times \mathbb{R}^M)$ and $u_\ell \rightarrow u$ in $C(0 \times \mathbb{R} \times \mathbb{R}^M)$. Then u is a viscosity solution of $F = 0$.

We prove this result to illustrate the nature of Definition 1 at least once. For proofs of subsequent results the reader is referred to [1].

Proof. Let $\varphi \in \mathcal{D}(0)^+$, $k \in \mathbb{R}$ and $\tilde{y} \in E_+(\varphi(u - k))$. Then $\varphi(\tilde{y})(u(\tilde{y}) - k) > 0$ and so $\varphi(\tilde{y})(u_\ell(\tilde{y}) - k) > 0$ for large ℓ . Thus $E_+(\varphi(u_\ell - k)) \neq \emptyset$ for large ℓ and, by assumption, there exists y_ℓ such that

$$\varphi(y)(u_\ell(y) - k) < \varphi(y_\ell)(u_\ell(y_\ell) - k) \quad \text{for} \quad y \in 0 \tag{1.5}$$

and

$$F_\ell(y_\ell, u_\ell(y_\ell), - \frac{(u_\ell(y_\ell) - k)}{\varphi(y_\ell)} D\varphi(y_\ell)) < 0. \tag{1.6}$$

Passing to a subsequence if necessary, we assume $y_\ell \rightarrow \hat{y} \in 0$. Taking the limits in (1.5) and (1.6) shows that $\hat{y} \in E_+(\varphi(u - k))$ and

$$F(\hat{y}, u(\hat{y}), - \frac{(u(\hat{y}) - k)}{\varphi(\hat{y})} D\varphi(\hat{y})) < 0.$$

Hence u is a viscosity solution of $F < 0$. Similarly, u is a viscosity solution of $F > 0$, and hence a viscosity solution of $F = 0$.

In the above proof we used the mild requirements of Definition 1 in that from $E_+(\varphi(u - k)) \neq \emptyset$ we concluded only the existence of some $\hat{y} \in E_+(\varphi(u - k))$ for which the desired inequality held. However, one can prove Definition 1 could

be strengthened a great deal without altering the notion defined. The next result outlines the extent to which this is true. For $\psi \in C(0)$ we let

$$d(\psi) = \{y \in O : D\psi(y) \text{ exists}\}$$

where " $D\psi(y)$ exists" means ψ is (Frechét) differentiable at y .

Theorem 2. Let $u, \varphi, \psi \in C(0)$ and u be a viscosity solution of $F = 0$. Then

$$F(y, u, -\frac{(u - \psi)}{\varphi} D\varphi + D\psi) \leq 0 \text{ everywhere on } E_+(\varphi(u - \psi)) \cap d(\varphi) \cap d(\psi)$$

and

$$F(y, u, -\frac{(u - \psi)}{\varphi} D\varphi + D\psi) \geq 0 \text{ everywhere on } E_-(\varphi(u - \psi)) \cap d(\varphi) \cap d(\psi).$$

Using Theorem 2 (and the proof of Theorem 2) one can prove:

Theorem 3. (Consistency). If u is a viscosity solution of $F = 0$, then

$$F(y, u(y), Du(y)) = 0 \text{ on } d(u).$$

In particular, a Lipschitz continuous viscosity solution of $F = 0$ satisfies the equation pointwise almost everywhere.

As a final example of the many results of [1, Section I], we consider the situation in which O is divided into two open parts O_- and O_+ by a surface Γ . Let $\vec{n}(y)$ be the normal to Γ at $y \in \Gamma$ and assume $\vec{n}(y)$ points into O_+ . Let $u \in C(0)$ be u_- in O_- and u_+ in O_+ and assume u_{\pm} is continuously differentiable in O_{\pm} respectively. Let u_{\pm} be classical solutions of $F = 0$ in O_{\pm} . We ask how the jump in Du across Γ is restricted in order that u be a viscosity solution of $F = 0$ in O (it is assumed that Du_{\pm} extend continuously to Γ). The answer is that u is a viscosity solution in O if and only if:

$$\text{For every } y \in \Gamma \text{ and } \xi \in I \quad (1.7)$$

$$F(y, u(y), p_{\Gamma} Du_{\pm}(y) + \xi \vec{n}(y))(\xi - Du_-(y) \cdot \vec{n}(y)) \geq 0$$

where I is the interval with endpoints $Du_+(y) \cdot \vec{n}(y)$ and $Du_-(y) \cdot \vec{n}(y)$, $a \cdot b$ means the inner-product of a and b , and $p_{\Gamma} Du_{\pm}(y)$ means the (equal) projection of $Du_+(y)$ and $Du_-(y)$ on the tangent space to Γ at y . This criterion rules out the solution $\max(0, t - |x|)$ of (0.7) in the introduction and all solutions of (1.4) mentioned in the text except $1 - |x|$, as the reader can verify.

SECTION 2. MODEL EXISTENCE AND UNIQUENESS RESULTS

We now indicate, in model problems, some of the very general statements concerning existence and uniqueness of solutions of (DP) and (CP) which are valid for viscosity solutions. Let us consider the Cauchy problem

$$u_t + H(Du) = 0, \quad t > 0, x \in \mathbb{R}^N$$

$$u(x, 0) = u_0(x) \quad (2.1)$$

where $H \in C(\mathbb{R}^N)$. Let $BUC(\Omega)$ denote the bounded and uniformly continuous functions on Ω . We have:

Theorem 4. Let $u_0 \in BUC(\mathbb{R}^N)$. Then there is exactly one function $u \in C(\mathbb{R}^N \times [0, \infty))$ with the following properties:

- (i) u is a viscosity solution of $u_t + H(Du) = 0$ on $\mathbb{R}^N \times (0, \infty)$.
- (ii) $u(x, 0) = u_0(x)$ for $x \in \mathbb{R}^N$.
- (iii) $u \in BUC(\mathbb{R}^N \times [0, T])$ for each $T > 0$.

This result is completely new. If, e.g., $H \equiv 0$ and u_0 is nowhere differentiable, then $u(x, t) \equiv u_0(x)$ is nowhere differentiable in x . The ability to formulate an existence and uniqueness result for (2.1) in our generality thus requires the ability to discuss nondifferentiable solutions of the equation, and this has not been possible before. However, the existence assertions can be proved by expanding on the arguments in the introduction concerning the relationship of the vanishing viscosity method and the notion of viscosity solutions, so we can adapt known methods here. The uniqueness is then the main new point. Theorem 4 allows one to define a family of mappings

$S(t) : BUC(\mathbb{R}^N) \rightarrow BUC(\mathbb{R}^N)$ by letting $S(t)u_0(\cdot)$ be $u(\cdot, t)$ where u is the solution of (2.1). One can also show that for $t, \tau > 0$ and $u_0, v_0 \in BUC(\mathbb{R}^N)$

$$S(t)S(\tau) = S(t + \tau) ,$$

$$\|S(t)u_0 - S(t)v_0\|_{L^\infty(\mathbb{R}^N)}^+ \leq \|u_0 - v_0\|_{L^\infty(\mathbb{R}^N)}^+ ,$$

$$|S(t)u_0(x + y) - S(t)u_0(x)| \leq \sup_{z \in \mathbb{R}^N} |u_0(z + y) - u_0(z)| .$$

The first two properties show that $S(t)$ is a semigroup of order-preserving nonexpansive mappings on $BUC(\mathbb{R}^N)$.

The more complex problem

$$\begin{aligned} u_t + H(x, t, u, Du) &= 0 , \\ u(x, 0) &= u_0(x) \end{aligned} \tag{2.2}$$

requires restrictions on H to have the validity of the uniqueness assertion. For example, if

- (i) $H(x, t, r, p)$ is nondecreasing in r ,
- (ii) $H(x, t, r, p)$ is uniformly continuous on $\mathbb{R}^N \times [0, T] \times [-R, R] \times \{p : |p| \leq R\}$ for each $R > 0$,

and

- (iii) $\lim_{\alpha \rightarrow 0} \sup \{ |H(x, t, r, p) - H(y, t, r, p)| : |x - y| \leq \alpha, |x - y||p| \leq R, 0 \leq t \leq T, |r| \leq R \} = 0$
for each $R > 0$,

then bounded viscosity solutions of (2.2) which assume their initial-values uniformly are unique. Note that (2.3) (i) rules out conservation law form for (2.2). Assumptions on H may be weakened in exchange for stricter requirements

on solutions. See [1]. The necessity of some requirement like (2.3) (iii) can be seen by considering the linear Hamiltonian $H(x, t, u, p) = b(x) \cdot p$, in which case (2.3) (iii) reduces to Lipschitz continuity of $b(x)$.

From classical considerations one expects problems like (2.2) to exhibit finite propagation speeds under suitable assumptions. This still holds true in the generality of viscosity solutions. For example, we have:

Theorem 5. Let H satisfy (2.3) (i), $L, R, T > 0$ and u, v be viscosity solutions of $u_t + H(x, t, u, Du) = 0$ on $Q_T = \mathbb{R}^N \times [0, T]$. Assume that

$$C = \max(\|Du\|_{L^\infty(Q_T)}, \|Dv\|_{L^\infty(Q_T)}), \quad m = \max(\|u\|_{L^\infty(Q_T)}, \|v\|_{L^\infty(Q_T)})$$

are finite and

$$|H(x, t, r, p) - H(x, t, r, q)| \leq L|p - q| \quad \text{for } |p|, |q| \leq C, |r| \leq m,$$

$$|x| \leq R - Lt \quad \text{and} \quad 0 \leq t \leq T.$$

Then $u(x, 0) \leq v(x, 0)$ on $|x| \leq R$ implies $u \leq v$ on $|x| \leq R - Lt, 0 \leq t \leq T$.

Of course, no such result holds if H is merely continuous in p .

While we have, so far, focused on the pure Cauchy problem, the results of [1] cover both (CP) and (DP). Moreover, they are more precise and general in the cases we have considered. We conclude by giving a uniqueness and continuous dependence proof in the simplest situation to illustrate the basic argument without technical complexities. Let $H \in C(\mathbb{R}^N)$, and

$$u, v, m, n \in C_0(\mathbb{R}^N) \quad (2.4)$$

where $C_0(\mathbb{R}^N)$ denotes the continuous functions on \mathbb{R}^N vanishing at infinity.

Assume that

$$\begin{aligned} u + H(Du) &\leq m, \\ v + H(Dv) &\geq n, \end{aligned} \quad (2.5)$$

in the viscosity sense. We prove that

$$u(x) - v(x) \leq \|(m - n)^+\|_{L^\infty(\mathbb{R}^N)} \quad \text{for } x \in \mathbb{R}^N. \quad (2.6)$$

We may assume that

$$M = \max_{\mathbb{R}^N} (u(x) - v(x)) > 0.$$

Choose $\varphi \in \mathcal{D}(\mathbb{R}^N)^+$ such that

$$\varphi(0) = 1, \quad 0 \leq \varphi \leq 1 \quad \text{and} \quad \varphi(x) = 0 \quad \text{for } |x| > 1. \quad (2.7)$$

Set $\varphi_\alpha(x) = \varphi(x/\alpha)$ and let $x_\alpha, y_\alpha \in \mathbb{R}^N$ satisfy

$$M_\alpha = \varphi_\alpha(x_\alpha - y_\alpha)(u(x_\alpha) - v(y_\alpha)) \geq \varphi_\alpha(x - y)(u(x) - v(y)) \quad \text{for } x, y \in \mathbb{R}^N. \quad (2.8)$$

The existence of x_α, y_α follows from the assumptions that $u, v \rightarrow 0$ at ∞ and φ_α has compact support. Putting $y = x$ in (2.8) and using $\varphi_\alpha \leq 1$ we find

$$M < M_\alpha < u(x_\alpha) - v(y_\alpha) . \quad (2.9)$$

Since $x_\alpha \in E_+(\varphi(\cdot - y_\alpha)(u(\cdot) - v(y_\alpha)))$ and (2.5) holds, Theorem 2 gives

$$u(x_\alpha) + H(-\frac{(u(x_\alpha) - v(y_\alpha))}{\varphi_\alpha(x_\alpha - y_\alpha)} D\varphi_\alpha(x_\alpha - y_\alpha)) < m(x_\alpha) . \quad (2.10)$$

Similarly,

$$v(y_\alpha) + H(-\frac{(u(x_\alpha) - v(y_\alpha))}{\varphi_\alpha(x_\alpha - y_\alpha)} D\varphi_\alpha(x_\alpha - y_\alpha)) > n(y_\alpha) \quad (2.11)$$

where we used that $(D\varphi)(x - y) = -D_y\varphi(x - y)$. Subtracting (2.11) from (2.10) yields

$$u(x_\alpha) - v(y_\alpha) < m(x_\alpha) - n(y_\alpha) = m(x_\alpha) - n(x_\alpha) + n(x_\alpha) - n(y_\alpha) . \quad (2.12)$$

Now $|x_\alpha - y_\alpha| < \alpha$ since $\varphi_\alpha(x_\alpha - y_\alpha) > 0$. Thus $|n(x_\alpha) - n(y_\alpha)| < \rho_n(\alpha)$ where ρ_n is the modulus of continuity of n . This with (2.12) and (2.9) give

$$\max_{\mathbb{R}^N} (u - v) < \|m - n\|_{L^\infty(\mathbb{R}^N)} + \rho_n(\alpha)$$

and the result follows upon letting $\alpha \rightarrow 0$.

REFERENCES

1. Crandall, M. G. and Lions, P. L., Viscosity solutions of Hamilton-Jacobi equations, Mathematics Research Center Technical Summary Report #2259, University of Wisconsin-Madison, 1981.
2. Crandall, M. G. and Lions, P. L., Approximation of viscosity solutions of Hamilton-Jacobi equations, in preparation.
3. Fleming, W. H., Nonlinear partial differential equations - Probabilistic and game theoretic methods, in: Problems in Nonlinear Analysis, CIME, Ed. Cremonese, Rome, (1971).
4. Friedman, A., The Cauchy problem for first order partial differential equations, Indiana Univ. Math. J. 23 (1973) 27-40.
5. Kružkov, S. N., First order quasilinear equations with several space variables, Math. USSR-Sb. 10 (1970) 217-243.
6. Lions, P. L., Generalized solutions of Hamilton-Jacobi equations, to appear.
7. Vol'pert, A. I., The spaces BV and quasilinear equations, Math. USSR-Sb. 2 (1967) 225-267.

This Page Intentionally Left Blank

BIFURCATION OF STATIONARY VORTEX CONFIGURATIONS

Julian I. Palmore

University of Illinois
Department of Mathematics
1409 West Green Street
Urbana, Illinois 61801 U.S.A.

We examine the dynamics problem of finitely many vortices in a circular disk and study stationary configurations of the vortices. For any fixed number of two or more vortices we prove that there are families of stationary vortex configurations in which bifurcation occurs. Sharp lower bounds on the numbers of stationary configurations are obtained by topological methods.

INTRODUCTION

An old problem of the dynamics of finitely many vortices in the plane is restated for vortices interacting in a circular disk. Our purpose here is to demonstrate the way in which topological and analytical methods can be applied to the problem of vortices in a disk to yield new sharp results on the bifurcation of solutions within families of stationary configurations. We examine the vortex problem as a Hamiltonian dynamical system and use critical point theory to study the stationary configurations.

Recall that a stationary configuration of vortices is a configuration such that each vortex moves uniformly in a circle so as to maintain the same relative configuration. Thus, there is a uniformly rotating coordinate system in which the configuration of vortices is fixed.

We begin the discussion by stating a main result on bifurcation within families of stationary configurations.

Theorem A. For any $n \geq 2$ and for any choice of positive circulations $\kappa_1, \dots, \kappa_n$, for vortices in a disk, there are $n!/2$ families of stationary configurations in which bifurcations occur.

In order to prove Theorem A we use a principle which is well-known in critical point theory. We select families of stationary configurations from the planar problem and continue these solutions into the problem of vortices in a disk. The index of each stationary configuration is known from the planar problem. By following the continuation of the solutions toward the boundary of the disk we prove that an index change occurs at a degeneracy within the family. The principle employed is that a degeneracy in a family of solutions accompanied by a change in the index implies bifurcation.

The interplay of ideas from celestial mechanics and dynamical systems is evident in the study of the vortex problem. Compare the methods and results of [1]. In [2] we announced the idea of using topological methods, useful in studying celestial mechanics and the relative equilibria solutions, to study the vortex problem.

A second result is on the classification of stationary vortex configurations. In [3] and [4] we state several results on the classification of stationary configurations in Kirchhoff's problem. Here, the classification for the vortices in the plane carries over to this setting and is modified where bifurcations occur.

As a starting point for the proof of Theorem A we state the following theorem on the existence of particular stationary configurations.

Theorem B. For any $n \geq 2$ and for any choice of positive circulations $\kappa_1, \dots, \kappa_n$, for vortices in a disk, there exist $n!/2$ families of collinear stationary configurations.

STATIONARY VORTEX CONFIGURATIONS AND THE HAMILTONIAN

We identify stationary vortex configurations with critical points of the restricted Hamiltonian. Let E^2 denote the Euclidean plane with inner product \langle, \rangle and norm $\| \cdot \|$. We write all functions and their derivatives using vector notation for the convenience of the analysis.

Let $D^2 \subset E^2$ denote the open unit disk, $D^2 = \{x \in E^2 \mid \|x\| < 1\}$. The configuration space of the n -vortex problem in the disk with positive circulations $\kappa_1, \dots, \kappa_n$ is given by

$$M = (D^2)^n - \Delta$$

where the diagonal Δ is the set of all configurations such that two or more vortices coincide. Thus, $\Delta = \bigcup_{i < j} \Delta_{ij}$ and $\Delta_{ij} = \{(x_1, \dots, x_n) \in (D^2)^n \mid x_i = x_j\}$. Here the union is over all $i < j$.

The Hamiltonian function H is defined on M ; thus, M is also the phase space of the problem. Hamilton's equations are not used directly; we write the Hamiltonian in vector notation as

$$H(x_1, \dots, x_n) = - \sum_{i \neq j} \kappa_i \kappa_j \log \|x_i - x_j\| \\ + \frac{1}{2} \sum_{1 \leq i, j \leq n} \kappa_i \kappa_j \log((1 - \|x_i\|^2)(1 - \|x_j\|^2) + \|x_i - x_j\|^2).$$

Singularities in H arise on the boundary of the disk and on the diagonal. Those on the boundary appear in the terms in the second sum for which $i = j$. Those singularities on the diagonal appear in the first sum.

The Hamiltonian $H(x)$ is not translation invariant. Therefore, the center of vorticity need not be constant.

The Hamiltonian is invariant under rotations of E^2 . Thus, the function $I : M \rightarrow \mathbb{R}$ defined by $I(x) = \frac{1}{2} \sum \kappa_i \|x_i\|^2$ is an integral of Hamilton's equation. Let S_α denote the set

$$S_\alpha = \{(x_1, \dots, x_n) \in (D^2)^n \mid I(x) = \alpha\}.$$

Then $S_\alpha \cap M$ is invariant under the flow defined by the dynamical system. As is the case for Kirchhoff's problem, for positive circulations of the vortices, collisions between vortices in a disk cannot occur for α sufficiently small.

Let H_α denote the restriction of H to S_α .

Theorem 1. A configuration $(x_1, \dots, x_n) \in S_\alpha$ is a stationary configuration of the vortices if and only if (x_1, \dots, x_n) is a critical point of H_α .

This statement gives us a critical point criterion by which bifurcation within families of stationary configurations can be detected.

DETECTING BIFURCATION

In order to detect bifurcation in a family of stationary configurations of vortices we examine the index of critical points as an analytic criterion. As an example of a one parameter family of critical points we consider the function $f_t: \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f_t(x, y) = tx^2 + (1-t)x^4 + y^2$. The derivative of f_t is $df_t(x, y) = (2tx + 4(1-t)x^3, 2y)$. For all $t \in \mathbb{R}$ $(x, y) = (0, 0)$ is a critical point. For $t \geq 0$, $(0, 0)$ is the only critical point. For $t < 0$, there are three critical points: $(0, 0)$, $((-t/2(1-t))^{1/2}, 0)$ and $(-(-t/2(1-t))^{1/2}, 0)$. The second derivative is $d^2f_t(x, y) = \text{diag}(2t + 12(1-t)x^2, 2)$. For $t \geq 0$ the origin is a minimum of f_t , nondegenerate for $t > 0$. By observing the index of the origin as a critical point of f_t , $t \in \mathbb{R}$, we observe that the index changes from 0 ($t \geq 0$) to 1 ($t < 0$). The index of a fixed neighborhood of the origin must be preserved when all critical points are nondegenerate. Thus, for $t < 0$ there must be at least two additional critical points in a neighborhood of the origin provided that f_t has nondegenerate critical points.

The procedure used above to detect a change in the critical points of f_t is the one we use to detect bifurcation within families of stationary configurations.

As an example of detecting bifurcation within a family of stationary configurations we consider the two vortex problem with unit circulation. We observe that the symmetric configuration is always a stationary configuration. The vortices are located on a line at positions $x_1 = (-\alpha, 0)$ and $x_2 = (\alpha, 0)$ for $0 < \alpha < 1$. The second derivative of H_α evaluated in the transverse translation direction $v_i = (0, 1)$, $i = 1, 2$ is

$$D^2H_\alpha(x)(v, v) = \frac{2}{\alpha^2} \left(\frac{1-\alpha^2}{1+\alpha^2} \right)^2 > 0$$

for all $0 < \alpha < 1$ and $v = (v_1, v_2)$. Therefore the transverse translation direction is not a null direction.

In the longitudinal transverse direction $v = (v_1, v_2)$, $v_i = (1, 0)$, $i = 1, 2$, the second derivative is evaluated as

$$D^2H_\alpha(x)(v, v) = -\frac{4}{1+\alpha^2} - \frac{4}{1-\alpha^2} - \frac{8\alpha^2}{(1-\alpha^2)^2} + 2\lambda$$

where $2\lambda = 2/\alpha^2 (1+3\alpha^4)/(1-\alpha^4)$. The value of α^2 for which the derivative is zero is the positive root of the polynomial equation $p(a) = 1 - 5a + 3a^2 - 7a^3 = 0$. We find $a = \alpha^2 = .2137403$ and $\alpha = .4623206 = \alpha^*$.

The two translation directions are directions in which the Hessian can be decomposed. As the size of the configuration is increased by increasing α , we find that for $\alpha < \alpha^*$, the Hessian $D^2H_\alpha(x)(v, v)$ is positive so that $D^2H_\alpha(x)$ is positive definite for $\alpha < \alpha^*$. For $\alpha = \alpha^*$, the Hessian is degenerate with rank equal to 1. For $\alpha > \alpha^*$, the stationary configuration is a saddle point of H_α . In order to maintain the consistency of the topological type of a fixed neighborhood of the symmetric stationary configuration, we must have two additional critical points, both of them minima with index 0. These critical points are the translates in the direction v of the symmetric configuration.

MAIN THEOREMS

In this section we state two results and prove that degeneracy in a translation mode always occurs along a family of collinear vortex configurations. Thus, we prove Theorem A.

Several results are needed. Compare [4].

Theorem 2. For any $n \geq 2$ and for any choice of positive circulations $\kappa_1, \dots, \kappa_n$, for the vortex problem in the plane, there are $n!/2$ collinear stationary configurations. As critical points of the reduced Hamiltonian \tilde{H} these configurations are nondegenerate with index equal to $n - 2$.

Here \tilde{H} denotes the Hamiltonian function for Kirchhoff's problem induced on complex projective space $\mathbb{P}_{n-2}(\mathbb{C}) - \Delta_{n-2}$ which arises as the reduced phase space in the vortex problem as well as the reduced configuration space of relative equilibria in the planar n -body problem of celestial mechanics. Compare [1].

The index of a critical point $x \in S_\alpha$ is the maximal dimension of a subspace of the tangent space of S_α on which the Hessian $D^2H_\alpha(x)$ is negative definite.

Theorem 2 enables us to prove Theorem B by using the implicit function theorem and the observation that for the n -vortex problem in the disk those stationary configurations for α sufficiently small must approximate stationary vortex configurations in the plane.

Let T_x denote the tangent space of $S_\alpha \cap (E^1)^n$ at $x \in S_\alpha$ and let T_x^\perp denote the normal space. The Hessian $D^2H_\alpha(x)$ splits on $T_x S_\alpha$ as $D^2H_\alpha(x) = D^2H_\alpha(x)|_{T_x} \oplus D^2H_\alpha(x)|_{T_x^\perp}$.

Let $v = (1, \dots, 1) \in (E^1)^n$ and $v = (1^\perp, \dots, 1^\perp) \in (E_1^\perp)^n$ denote the longitudinal and transverse translation directions, respectively.

Theorem 3. Let $(x_1, \dots, x_n) \in (E^1)^n \cap S_\alpha$ be a collinear stationary configuration of the vortices with positive circulations $\kappa_1, \dots, \kappa_n$. The Hessian $D^2\tilde{H}_\alpha(x)$ is non null on the normal subspace of the tangent space.

As in the example of two vortices we search for bifurcations within the collinear configurations in the longitudinal translation direction.

Let $F_\alpha(x_1, \dots, x_n)$ denote the family of collinear stationary configurations, parameterized by α , which contains $(x_1, \dots, x_n) \in (E^1)^n \cap S_\alpha$, a stationary vortex configuration.

Theorem 4. The Hessian $D^2\tilde{H}_\alpha(x_\alpha)$, $x_\alpha \in F_\alpha(x_1, \dots, x_n)$, is degenerate in the longitudinal translation direction for some $\alpha > 0$.

Proof. The Hessian $D^2\tilde{H}_\alpha(x_\alpha)(v, v)$ is evaluated as

$$D^2\tilde{H}_\alpha(x_\alpha)(v, v) = \sum_{1, j} \kappa_1 \kappa_j \left\{ \frac{-(1 - \|x_1\|^2) - (1 - \|x_j\|^2) + 4\langle x_1, x_j \rangle}{(1 - \|x_1\|^2)(1 - \|x_j\|^2) + \|x_1 - x_j\|^2} \right. \\ \left. - 2 \frac{(\langle x_1, 1 \rangle (1 - \|x_j\|^2) + \langle x_j, 1 \rangle (1 - \|x_1\|^2))}{((1 - \|x_1\|^2)(1 - \|x_j\|^2) + \|x_1 - x_j\|^2)^2} \right\} + (\sum \kappa_1) \lambda.$$

Here λ is the Lagrange multiplier determined by

$$\lambda = \frac{-D\tilde{H}_\alpha(x_\alpha)(x_\alpha)}{2I(x_\alpha)}.$$

Within the sum, the terms in $i = j$ contribute

$$\sum_{i=1}^n \frac{2\kappa_i^2}{1-\|x_i\|^2} \left(-\frac{1+\|x_i\|^2}{1-\|x_i\|^2} + \frac{\sum \kappa_i}{2I(x)} \right).$$

For α sufficiently large, this sum is negative as x_1 and x_n approach the boundary of the disk D^2 . The sum over $i \neq j$ reduces to

$$\begin{aligned} \sum_{i \neq j} \kappa_i \kappa_j & \frac{\{(-2+\|x_i\|^2 + \|x_j\|^2)((1-\|x_i\|^2)(1-\|x_j\|^2) + \|x_i - x_j\|^2) \\ & + 4(x_i, x_j)\|x_i - x_j\|^2 - 2(\|x_i\|^2 + \|x_j\|^2 - 2\|x_i\|^2\|x_j\|^2)\}}{((1-\|x_i\|^2)(1-\|x_j\|^2) + \|x_i - x_j\|^2)^2}. \end{aligned}$$

Let $\|x_i\| = \alpha$ and $\|x_j\| = \beta$ and examine the function

$$\begin{aligned} f(\alpha, \beta) &= (-2+\alpha^2+\beta^2)((1-\alpha^2)(1-\beta^2) + (\alpha-\beta)^2) \\ &+ 4\alpha\beta(\alpha-\beta)^2 - 2(\alpha^2+\beta^2-2\alpha^2\beta^2) \end{aligned}$$

on the unit square $0 \leq \alpha, \beta \leq 1$. The only zero of $f(\alpha, \beta)$ is at $(\alpha, \beta) = (1, 1)$. Thus, there are no zeros of f in the interior and $f(\alpha, \beta) < 0$ for $0 < \alpha, \beta < 1$. Consequently, each term in the sum over $i \neq j$ is negative. Our analysis shows that for any collinear stationary configuration the first sum is negative. As $\|x_1\|$ or $\|x_n\|$ approaches 1 the Hessian is eventually negative. The index of the critical point has changed by 1 (at least) along the family of stationary configurations.

ACKNOWLEDGEMENT

This research was supported in part by a grant from the National Science Foundation, U.S.A.

REFERENCES

- [1] J. Palmore, Measure of degenerate relative equilibria, I, *Annals of Math.* **104** (1976), p. 421-429.
- [2] J. Palmore, Relative equilibria of interacting vortices in two dimensions, Abstract 768-58-4, *Notices of the Amer. Math. Soc.* (1979), p. A-482.
- [3] J. Palmore, Central configurations, CW complexes and the homology of projective spaces, in *Classical Mechanics and Dynamical Systems*, ed. by R. Devaney and Z. Nitecki, Marcel-Dekker, NY (1981), p. 225-237.
- [4] J. Palmore, Relative equilibria of interacting vortices in two dimensions (1981), to appear in *Proc. Nat. Acad. of Sciences*.

This Page Intentionally Left Blank

EXACT INVARIANTS FOR TIME-DEPENDENT
NONLINEAR HAMILTONIAN SYSTEMS

H. R. Lewis

Los Alamos National Laboratory
P. O. Box 1663
Los Alamos, NM 87545
U.S.A.

and

P. G. L. Leach

Department of Applied Mathematics
La Trobe University
Bundoora 3083
Australia

Two methods, one based on canonical transformations and one based on an assumed structure, are used to determine exact invariants for certain Hamiltonians of the type $H = \frac{1}{2} p^2 + V(q,t)$. Invariants are found explicitly for a class of nonlinear, time-dependent potentials $V(q,t)$. The former method is then developed to find exact invariants for Hamiltonians of the form $H = H(q,p,\rho(t),\beta(t))$.

I. INTRODUCTION

The search for exact invariants for nonautonomous Hamiltonian systems has been prompted by theoretical studies in plasma physics and quantum mechanics where such systems play an important role. Apart from applications in these and other fields, our results are also of interest in the theory of canonical transformations in analytical dynamics.

Exact invariants for nonautonomous linear systems have been found during recent years by using a variety of methods [1], and invariants have also been found for some nonlinear systems [2]. The work that is reported briefly here has as its aim the determination of exact invariants for much broader classes of nonautonomous systems. In Section II, we outline a method that uses canonical transformations for Hamiltonians of the form

$$H = \frac{1}{2} p^2 + V(q,t) \tag{1.1}$$

to obtain invariants for a certain class of potentials. The contents of Section II are described in detail in a forthcoming article [3]. The invariants obtained in Section II are all quadratic in p . In Section III, we begin with the ansatz that an invariant is quadratic in p ,

$$I = f_0(q,t) + pf_1(q,t) + p^2f_2(q,t) \quad , \quad (1.2)$$

and solve the equation

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + [I, H] = 0 \quad (1.3)$$

directly, where H has the form (1.1). The result is invariants for a wider class of potentials than in Section II. In Section IV we return to the method of Section II, but do not restrict the dependence of the Hamiltonian on p to be as in (1.1). The details of Sections III and IV will be published elsewhere [4,5].

II. CANONICAL TRANSFORMATION APPROACH

In the conventional treatment of canonical transformations, the generating function contains a mixture of new and old variables. For this work we employ an unconventional type of generating function, which is a function of the old variables only. Consider a canonical transformation from (q,p) to (Q,P) where

$$Q = Q(q,p,t) \quad , \quad P = P(q,p,t) \quad . \quad (2.1)$$

The original Hamiltonian $H(q,p,t)$ and the new Hamiltonian $K(Q(q,p,t), P(q,p,t), t)$ are related by

$$p \frac{dq}{dt} - H = P \frac{dQ}{dt} - K + \frac{dF}{dt} \quad , \quad (2.2)$$

where $F(q,p,t)$ is the generating function of the transformation. Treating all functions in (2.2) as functions of the old coordinates only, we separate the coefficients of $\frac{dq}{dt}$ and $\frac{dp}{dt}$ and the part that does not involve either to obtain

$$p - P \frac{\partial Q}{\partial q} - \frac{\partial F}{\partial q} = 0 \quad , \quad (2.3)$$

$$P \frac{\partial Q}{\partial p} + \frac{\partial F}{\partial p} = 0 \quad , \quad (2.4)$$

$$H + P \frac{\partial Q}{\partial t} - K + \frac{\partial F}{\partial t} = 0 \quad . \quad (2.5)$$

We can eliminate $F(q,p,t)$ from (2.3)-(2.5) in two ways, viz., by differentiating (2.5) with respect to q and (2.3) with respect to t or by differentiating (2.5) with respect to p and (2.4) with respect to t . We obtain, respectively,

$$\frac{\partial K}{\partial q} = \frac{\partial H}{\partial q} + \frac{\partial P}{\partial q} \frac{\partial Q}{\partial t} - \frac{\partial P}{\partial t} \frac{\partial Q}{\partial q} \quad , \quad (2.6)$$

$$\frac{\partial K}{\partial p} = \frac{\partial H}{\partial p} + \frac{\partial P}{\partial p} \frac{\partial Q}{\partial t} - \frac{\partial P}{\partial t} \frac{\partial Q}{\partial p} \quad . \quad (2.7)$$

So far the discussion has been quite general. We now introduce some constraints that will enable a result to be obtained. We introduce a nontrivial auxiliary function, $\rho(t)$, which is such that $\dot{\rho} \neq 0$, and we assume that the time dependence of the transformation is expressed completely by dependence of Q and P on ρ and $\dot{\rho}$,

$$Q = Q(q,p,\rho,\dot{\rho}) \quad , \quad P = P(q,p,\rho,\dot{\rho}) \quad . \quad (2.8)$$

We take

$$H = \frac{1}{2} p^2 + V(q,t) \quad , \quad K = K(P,\rho) \quad , \quad (2.9)$$

and assume that $V(q,t)$ cannot be written with its time dependence expressed entirely through $\rho(t)$ and $\dot{\rho}(t)$. Because the new Hamiltonian, K , does not depend on Q , the new momentum, P , will be an invariant.

With these constraints, (2.6) and (2.7) become

$$\frac{\partial K}{\partial P} \frac{\partial P}{\partial q} = \frac{\partial V}{\partial q} + \dot{\rho}[P, Q]_{qp} + \dot{\rho}[P, Q]_{q\dot{\rho}} \quad , \quad (2.10)$$

$$\frac{\partial K}{\partial P} \frac{\partial P}{\partial \rho} = p + \dot{\rho}[P, Q]_{p\rho} + \dot{\rho}[P, Q]_{p\dot{\rho}} \quad , \quad (2.11)$$

in which the bracket $[P, Q]_{\alpha\beta}$ is defined as

$$[P, Q]_{\alpha\beta} = \frac{\partial P}{\partial \alpha} \frac{\partial Q}{\partial \beta} - \frac{\partial P}{\partial \beta} \frac{\partial Q}{\partial \alpha} \quad . \quad (2.12)$$

In addition, we use the Poisson bracket relation for Q and P and the requirement of consistency of the time evolution of Q in the two coordinate systems, i.e.,

$$[Q, P]_{qp} = 1 \quad , \quad (2.13)$$

$$\frac{\partial K}{\partial P} = p \frac{\partial Q}{\partial q} - \frac{\partial V}{\partial q} \frac{\partial Q}{\partial \rho} + \dot{\rho} \frac{\partial Q}{\partial \rho} + \dot{\rho} \frac{\partial Q}{\partial \dot{\rho}} \quad . \quad (2.14)$$

The latter equation is also a consequence of (2.10), (2.11), and (2.13). The analysis of (2.10), (2.11), (2.13), and (2.14) is lengthy and will be given in detail elsewhere [3]. It is based on examination of the equations to uncover restrictions of functional dependences of the unknown functions on the variables q, p, ρ , and $\dot{\rho}$. We can illustrate the type of argument by describing how the functional dependence of P on its arguments is restricted as an immediate consequence of (2.10) and (2.11). In (2.11), each quantity except $\dot{\rho}$ manifestly depends on t only through explicit dependence on $\rho(t)$ and $\dot{\rho}(t)$. Therefore, either $[P, Q]_{p\dot{\rho}}$ vanishes, or $\dot{\rho}$ is expressible completely in terms of ρ and $\dot{\rho}$, or both. If $\dot{\rho}$ were expressible completely in terms of ρ and $\dot{\rho}$, then (2.10) would require that the t dependence of $\frac{\partial V}{\partial q}$ be expressible solely in terms of ρ and $\dot{\rho}$; but we have assumed this not to be the case. Therefore, we must have

$$[P, Q]_{p\dot{\rho}} = 0 \quad . \quad (2.15)$$

Considering this as a first-order partial differential equation for P , we then find

$$P(q, p, \rho, \dot{\rho}) = \Gamma(Q, q, \rho) \quad , \quad (2.16)$$

Where Γ is an arbitrary function.

The result of carrying the analysis to a conclusion can be summarized as follows [3]. We have derived an explicit invariant for any potential of the form

$$V(q, t) = \frac{f_0(a, t)}{c_2 - c_1 a} \left(\frac{1}{2} c_1 q^2 - c_0 q \right) + \frac{W(u)}{(c_2 - c_1 a)^2} \quad , \quad (2.17)$$

where $a(t)$ is any particular solution of the differential equation

$$\ddot{a} = f_0(a, t) \quad , \quad (2.18)$$

f_0 is an arbitrary function of its arguments, c_0 , c_1 and c_2 are arbitrary constants such that $c_1^2 + c_2^2 = 1$, $W(u)$ is an arbitrary function of u , and u is defined by

$$u = \frac{q - c_0(c_1 + c_2 a)}{c_2 - c_1 a} \quad . \quad (2.19)$$

We have found a canonical transformation for which the new Hamiltonian is

$$K(P, \rho) = \frac{P}{(c_2 - c_1 a)^2} \quad . \quad (2.20)$$

The invariant is the new momentum, which is given by

$$P = \frac{1}{2} v^2 + W(u) \quad , \quad (2.21)$$

where

$$v = (c_2 - c_1 a)p + \dot{a}(c_1 q - c_0) \quad . \quad (2.22)$$

The new coordinate is

$$Q = \frac{v}{|v|} \int^u \frac{du'}{\{2(P - W(u'))\}^{1/2}} + T(P) \quad , \quad (2.23)$$

where T is an arbitrary function.

These results may be interpreted in terms of a transformation to action-angle variables under a generalized canonical transformation

$$(q, p, t) \rightarrow (Q, P, T) \quad ,$$

where Q and P are given by (2.23) and (2.21), respectively, and

$$T = \int^t (c_2 - c_1 a(t'))^{-2} dt' \quad . \quad (2.24)$$

The new Hamiltonian is then simply the invariant P .

III. INVARIANTS QUADRATIC IN THE MOMENTUM

Because the invariants found in Section II are all quadratic in the momentum p , it is natural to inquire whether the admissible class of potentials associated with those invariants contains all potentials for which there exists an invariant quadratic in p . Thus, we assume a Hamiltonian

$$H = \frac{1}{2} p^2 + V(q, t) \quad (3.1)$$

and make the ansatz that there exists an invariant of the form

$$I = f_0(q, t) + p f_1(q, t) + p^2 f_2(q, t) \quad . \quad (3.2)$$

That is, I is to satisfy

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + [I, H]_{qp} = 0. \quad (3.3)$$

Substitution of (3.1) and (3.2) into (3.3) gives the system of partial differential equations

$$\frac{\partial f_2}{\partial q} = 0, \quad (3.4)$$

$$\frac{\partial f_1}{\partial q} + \frac{\partial f_2}{\partial t} = 0, \quad (3.5)$$

$$\frac{\partial f_0}{\partial q} - 2f_2 \frac{\partial V}{\partial q} + \frac{\partial f_1}{\partial t} = 0, \quad (3.6)$$

$$-f_1 \frac{\partial V}{\partial q} + \frac{\partial f_0}{\partial t} = 0. \quad (3.7)$$

The solution of (3.4) - (3.7) is relatively straightforward and will be described elsewhere [4]. The result is the following. There exists an invariant quadratic in p if, and only if, the potential is of the form

$$V(q, t) = -F(t)q + \frac{1}{2}\Omega^2(t)q^2 + \frac{1}{\rho_1^2} W\left(\frac{q - \rho_2}{\rho_1}\right), \quad (3.8)$$

where F , Ω , and W are arbitrary functions, ρ_1 and ρ_2 are any particular solutions of

$$\ddot{\rho}_1 + \Omega^2(t)\rho_1 - \frac{k}{\rho_1^3} = 0, \quad (3.9)$$

$$\ddot{\rho}_2 + \Omega^2(t)\rho_2 = F(t) \quad , \quad (3.10)$$

and k is an arbitrary constant. The invariant quadratic in p is

$$I(q, p, t) = \frac{1}{2} [\rho_1(p - \dot{\rho}_2) - \dot{\rho}_1(q - \rho_2)]^2 + \frac{1}{2} k \left(\frac{q - \rho_2}{\rho_1} \right)^2 + W \left(\frac{q - \rho_2}{\rho_1} \right) \quad . \quad (3.11)$$

This result is a generalization of the result obtained in Section II because the potential may now contain two independent functions of time.

IV. FURTHER DEVELOPMENT OF THE CANONICAL TRANSFORMATION APPROACH

In Section II, our canonical transformation approach was applied to a Hamiltonian equal to $\frac{1}{2} p^2$ plus a potential whose time dependence was assumed not to be expressible entirely through dependence on $\rho(t)$ and $\dot{\rho}(t)$. We now assume that the time dependence of the Hamiltonian is expressible solely in terms of $\rho(t)$ and $\dot{\rho}(t)$, and we allow a general dependence of the Hamiltonian on p . That is, we take

$$H = H(q, p, \rho, \dot{\rho}) \quad . \quad (4.1)$$

The new canonical variables Q and P again are taken to be functions of q , p , ρ and $\dot{\rho}$. Equations (2.10) and (2.11) now read

$$\frac{\partial K}{\partial P} \frac{\partial P}{\partial q} = [H, p]_{qp} + \dot{\rho}[Q, P]_{\rho q} + \ddot{\rho}[Q, P]_{\dot{\rho} q} \quad , \quad (4.2)$$

$$\frac{\partial K}{\partial P} \frac{\partial P}{\partial p} = -[H, q]_{qp} - \dot{\rho}[Q, P]_{p\rho} - \ddot{\rho}[Q, P]_{p\dot{\rho}} \quad . \quad (4.3)$$

In (4.2) and (4.3), all quantities except $\ddot{\rho}$ depend on t only through dependence on ρ and $\dot{\rho}$. Therefore, we either take

$$[Q, P]_{\dot{\rho}q} = 0 \quad , \quad [Q, P]_{p\dot{\rho}} = 0 \quad , \quad (4.4)$$

or require ρ to satisfy a differential equation of the form

$$\dot{\rho} = g(\rho, \dot{\rho}) \quad . \quad (4.5)$$

(As before, we require $\dot{\rho} \neq 0$.) The former choice is too restrictive. Adopting the latter, we write

$$\dot{\rho}[Q, P]_{\dot{\rho}q} = f_1(q, p, \rho, \dot{\rho}) \quad , \quad \dot{\rho}[Q, P]_{p\dot{\rho}} = -f_2(q, p, \rho, \dot{\rho}) \quad . \quad (4.6)$$

These two equations may be combined to give a single homogeneous partial differential equation for P ,

$$[Q, P]_{\dot{\rho}q} f_2(q, p, \rho, \dot{\rho}) + [Q, P]_{p\dot{\rho}} f_1(q, p, \rho, \dot{\rho}) = 0 \quad . \quad (4.7)$$

The solution is

$$P(q, p, \rho, \dot{\rho}) = \Gamma\{Q(q, p, \rho, \dot{\rho}), u(q, p, \rho), \rho\} \quad (4.8)$$

where $u(q, p, \rho)$ is related to f_1 and f_2 by

$$f_1(q, p, \rho, \dot{\rho}) = f(q, p, \rho, \dot{\rho}) \frac{\partial u}{\partial q} \quad , \quad f_2(q, p, \rho, \dot{\rho}) = f(q, p, \rho, \dot{\rho}) \frac{\partial u}{\partial p} \quad , \quad (4.9)$$

and f is an arbitrary function.

Substituting Γ for P in (4.2) and (4.3), taking two combinations of the resulting equations, and rewriting the Poisson bracket requirement on Q and P , we obtain the fundamental equations

$$\frac{\partial K}{\partial P} = [Q, H] + \dot{\rho} \frac{\partial Q}{\partial \rho} + \dot{\rho} \frac{\partial Q}{\partial \dot{\rho}} \quad , \quad (4.10)$$

$$\frac{\partial K}{\partial P} \frac{\partial \Gamma}{\partial Q} = - \frac{\partial \Gamma}{\partial u} \dot{u} - \frac{\partial \Gamma}{\partial \rho} \dot{\rho} \quad , \quad (4.11)$$

$$\frac{\partial \Gamma}{\partial u} = \frac{1}{[Q, u]} \quad , \quad (4.12)$$

where

$$\dot{u} \equiv \dot{\rho} \frac{\partial u}{\partial \rho} + [u, H] \quad . \quad (4.13)$$

Detailed analysis of restrictions on functional dependences required by these equations finally leads to a result. The entire derivation will be published elsewhere [5]. Here we summarize the results.

We find that Q may be written as a function of two canonical variables, u (as defined above) and v :

$$Q(q, p, \rho, \dot{\rho}) = R(u, v) \quad , \quad (4.14)$$

where

$$v = I(q, p, \rho) - J(u, \rho, \dot{\rho}) \quad (4.15)$$

and

$$[u, v] = [u, I] = 1 \quad . \quad (4.16)$$

The functions u and v also satisfy

$$\frac{\partial \dot{u}}{\partial \dot{\rho}} + \frac{\partial J}{\partial \rho} [u, \dot{u}] = 0 \quad , \quad (4.17)$$

$$\frac{\partial \dot{v}}{\partial \dot{\rho}} + \frac{\partial J}{\partial \rho} [u, \dot{v}] = 0 \quad . \quad (4.18)$$

The dot denotes total time differentiation, as in (4.13), and a bracket without subscripts denotes the usual Poisson bracket with respect to q and p .

As in the problem considered in Section II, there is no loss of generality in taking $K(P, \rho)$ to be linear in P and, in fact, we take

$$\begin{aligned} K(P, \rho) &= \beta(\rho)P \\ &= \beta(\rho) N(u, v) \quad , \end{aligned} \quad (4.19)$$

where $N(u, v)$ is the invariant. The permissible form of the Hamiltonian is given by

$$H(q, p, \rho, \dot{\rho}) = \beta(\rho) N(u, v) + a_0(q, p, \rho) + \dot{\rho} a_1(q, p, \rho) + a_2(u, \rho, \dot{\rho}) \quad , \quad (4.20)$$

where $a_2(u, \rho, \dot{\rho})$ is not linear in $\dot{\rho}$. The functions a_0 , a_1 , and a_2 are related to u , v and $g(\rho, \dot{\rho}) (= \beta)$ by

$$[u, a_0] = 0 \quad , \quad \frac{\partial u}{\partial \rho} + [u, a_1] = 0 \quad , \quad (4.21)$$

$$[I, a_0] = G_0(u, \rho) \quad , \quad \frac{\partial I}{\partial \rho} + [I, a_1] = G_1(u, \rho) \quad , \quad (4.22)$$

$$g(\rho, \dot{\rho}) \frac{\partial J}{\partial \dot{\rho}} + \dot{\rho} \frac{\partial J}{\partial \rho} + \frac{\partial a_2}{\partial u} = G_0(u, \rho) + \dot{\rho} G_1(u, \rho) \quad . \quad (4.23)$$

Equations (4.21) - (4.23) together with the Poisson bracket relation for u and v constitute a consistent set of equations in themselves and are consistent with (4.17) and (4.18).

V. DISCUSSION

The results of the various analyses that have been described briefly here indicate that it is possible to find invariants for a wide class of time-dependent one-dimensional Hamiltonian systems. Such results will have generalizations to

higher dimensions as has been found by Lewis [6] for the three-dimensional counterpart of the problem described in Section III. The results reported in Section IV mark a fundamental change from those of Section II and III and earlier results. In the earlier work, canonical coordinates, here called (u,v) , have been obtained for which the transformation from (q,p) to (u,v) consisted in a coordinate transformation q to u with the momentum transformation from p to v being induced by the coordinate transformation. In Section IV, the possibility for more general transformations is admitted. Indeed, one might list the types of problem that could be treated in terms of the p dependence of u .

In Sections II and IV, the time dependence of the canonical transformations was through the single function $\rho(t)$ whereas in Section III we saw that two independent functions may arise. Part of our program of investigation of invariants for Hamiltonian systems is to consider, using the method of Sections II and IV, the effect of introducing several independent functions. The other part is to apply the method of Section IV to a Hamiltonian whose time dependence is not through $\rho(t)$ and $\dot{\rho}(t)$ alone. We hope to report on these matters soon.

ACKNOWLEDGMENTS

Much of the work described here was performed while one of us (PGLL) was a Visiting Scientist with the Center for Nonlinear Studies at the Los Alamos National Laboratory. Without the support of the Center and a travel grant from La Trobe University the visit would not have been possible.

REFERENCES

- [1] Lewis, H. R., Phys. Rev. Lett. 18 (1967) 510-512, and Euratom, Phys. Rev. Lett. 18 (1967) 636; Lewis, H. R., J. Math. Phys. 9 (1968) 1976-1986; Lewis, H. R., Phys. Rev. 172 (1968) 1313-1315; Sarlet, W. and Bahar, L. Y., Quadratic integrals for linear nonconservative systems and their connection with the inverse problem of Lagrangian dynamics (preprint, RUGsam 80-16, September 1980, Instituut voor Theoretische Mechanica, Rijksuniversiteit, Gent, B-9000 Gent, Belgium); Leach, P. G. L., J. Austral. Math. Soc. 20B (1977) 97-105; Leach, P. G. L., J. Math. Phys. 21 (1980) 300-304.
- [2] Sarlet, W. and Bahar, L. Y., Int. J. Nonlinear Mech. 15 (1980) 133-146; Leach, P. G. L., to be published in J. Math. Phys.; Ray, J. R. and Reid, J. L., J. Math. Phys. 20 (1979) 2054-2057; Ray, J. R., Phys. Lett. 78A (1980) 4-6.
- [3] Lewis, H. R. and Leach, P. G. L., Exact invariants for a class of time-dependent nonlinear Hamiltonian systems, to be published in J. Math. Phys.
- [4] Lewis, H. R. and Leach, P. G. L., A direct approach to finding exact invariants for one-dimensional time-dependent classical Hamiltonians, to be submitted to J. Math. Phys.
- [5] Leach, P. G. L., Lewis, H. R. and Sarlet, W., Exact invariants for a class of time-dependent nonlinear Hamiltonian Systems, III, to be submitted to J. Math. Phys.
- [6] Lewis, H. R., Invariants quadratic in the momentum for three-dimensional scalar and vector potentials, to be submitted to J. Math. Phys.

This Page Intentionally Left Blank

ISOLATING INTEGRALS IN GALACTIC DYNAMICS AND THE CHARACTER OF STELLAR ORBITS

William I. Newman

Department of Earth and Space Sciences
University of California
Los Angeles, California 90024

The necessary and sufficient conditions for the existence of separable solutions to the Hamilton-Jacobi equation are well known. We show how the Hamilton-Lagrange equations of motion can be employed to obtain the same integrability conditions and isolating integrals of motion. In particular, for potentials that are used in galactic dynamics that do not possess three isolating integrals, this approach facilitates the use of perturbation methods and the understanding of the so-called box, tube and shell orbits that have been observed in numerical studies.

INTRODUCTION AND REVIEW

Galaxies are ensembles of $\sim 10^{11}$ self-gravitating stars. Individual stellar orbits are essentially free from strong binary interactions (i.e. collision times are substantially longer than the age of the universe for typical galaxies). Indeed, each star appears to move under the influence of a gravitational potential which describes the cumulative effect of the mass distribution of the galaxy. The time scale for an orbit is $\sim 10^8$ years and a given star will have executed ~ 100 orbits moving at a speed of several hundred km/sec during the galaxy's existence. Many galaxies show little sign of evolution and, we can surmise, the gravitational potential associated with their mass distributions changes very little with time. In its most primitive form, galactic dynamics seeks to explain the qualitative character of stellar orbits due to the galactic gravitational potential. Then, we can hope to better understand how stellar orbits contribute to the stability of a galaxy to gravitational collapse and to the morphology of their host.

Since the stars which constitute a galaxy are so numerous and manifest only collective effects, we treat the galactic mass distribution as a fluid with mass density $\rho(x)$ at spatial position x . [For the reasons given above, we assume that ρ is independent of time.] Corresponding to this mass density is the gravitational potential $V(x)$ which satisfies the gravitational Poisson's equation

$$\nabla^2 V(x) = 4\pi G \rho(x) \quad . \quad (1)$$

From observations of galaxies, we are able to (approximately) deduce the mass density. In particular, we will consider the so-called "elliptical galaxies," for which surfaces of equal mass density are well-described by ellipsoids with unequal axes. By employing the Green's function for Poisson's equation, the solution for the potential in (1) reduces to a quadrature. The resulting potential is generally quite asymmetric but is otherwise smooth and free from oscillations. We wish, then, to understand the character of stellar orbits under the influence of this observationally-deduced potential.

In a conservative force field, the motion of a particle (i.e. a star) has at least one integral of motion, the energy of the particle. If this were the only integral

of motion, the star would be able to visit all regions of phase space that are energetically available to it given sufficient time. This "ergodic hypothesis" together with the random character of the resulting stellar orbits could undermine the global stability of a galaxy. The existence of other integrals of motion would reduce the region of phase space accessible to a star and eliminate the stochastic behavior that would otherwise result. Should extra integrals of motion exist, it would be possible to assemble a linear combination of stellar orbits which would reproduce the underlying mass density distribution, i.e. obtain a self-consistent description of dynamical equilibrium with a density profile corresponding to elliptical galaxies. Thus, we would like to determine the number (and nature) of integrals of motion for a given gravitational potential.

Loosely speaking, integrals of motion correspond to a quantity that is conserved in each of its coordinates. For example, a three-dimensional simple harmonic oscillator preserves the energy of vibration in each of its (principal) coordinate axes. In addition, if the "spring constant" is isotropic (i.e. the restoring force depends only upon displacement from equilibrium and not upon the direction of displacement), the frequency of vibration along each of the coordinate axes is the same, which then contributes two more integrals of the motion. From the six Lagrange or Hamilton equations of motion for a particle in a time-independent, conservative force field, we can have at most five integrals of motion. However, only the isotropic harmonic oscillator (described above) and the Kepler problem of celestial mechanics possess five integrals of motion (Goldstein, 1980). In those two cases, the fourth and fifth integrals of motion are a byproduct of the special symmetry and frequency degeneracy in the problem. Spherically symmetric potentials possess, in general, four integrals of motion, the energy and the three components of the vector angular momentum. (The fourth integral of motion can again be considered to be the result of a frequency degeneracy between the θ and ϕ motions of the particle.) In the absence of spherical symmetry, a maximum of three constants of the motion can be obtained.

The problem of determining how many integrals of motion exist for a given potential remains unsolved. In 1891, Stäckel determined the conditions under which the solution to Hamilton-Jacobi equation were additively separable, a situation wherein three integrals of the motion naturally emerge. In this case, the dynamical behavior in each of the three coordinates would be periodic with a well-defined frequency. Usually, the three frequencies would be distinct and the behavior in each degree of freedom would be independent of the other two. The integrals of motion would describe the amount of energy or momentum that could be associated with each coordinate. Since the motion in each coordinate is periodic, it is customary to express each integral as an "action variable," a quantity that has the dimensions of angular momentum (or of energy times the period). It should be pointed out that each action variable establishes the frequency of motion along the corresponding coordinate as well as the spatial extent of the oscillation. In that sense, the integrals of motion determine the "classical turning points" along each of the independent axes. Clearly, this kind of coordinate separability is highly geometry-dependent. Stäckel (1891) showed that the necessary and sufficient conditions for separability (in the Hamilton-Jacobi problem) required that the potential possess certain symmetries in coordinate bases derived from confocal quadrics (Morse and Feshbach, 1953; Pars, 1965), a set of eleven coordinate systems obtained from conic sections which find frequent application in mathematical physics. This problem became fashionable in the 1920's and 1930's as physicists sought to understand the connection between classical and quantum mechanics. Eisenhart (1934) showed that the only orthogonal systems of coordinates for which separability of Schrödinger's equation was possible were those in which the Hamilton-Jacobi equation was separable. The transition from classical to quantum mechanics was further reinforced by the observation that the action variables were "quantized" in the process; i.e. they became eigenvalues of a separable set of ordinary differential equations. Lynden-Bell (1962) revived the issue of integrals of the motion in application to stellar (i.e. galactic) dynamics. Calling the constants of the motion "isolating

integrals," he proceeded to obtain integrability conditions that were quite similar to Stäckel's (1891) for orthogonal coordinate systems. These conditions, however, are not exclusive as the discovery by Hénon (see Berry, 1978) of a third integral for the Toda potential has shown. Indeed, the overthrow of the ergodic hypothesis (Kolmogorov, 1954) and the emergence of the Kolmogorov-Arnold-Moser theorem has revolutionized our way of thinking about integrability and galactic dynamics.

Elliptical galaxies may be considered to be spherically-symmetric objects which have experienced some asymmetrizing influence. In particular, we may regard the potential $V(x)$ as being a spherically-symmetric term with an additive aspherical perturbative potential. Thus, in the sense of Kolmogorov, Arnold and Moser, we might expect the perturbation to preserve three integrals of motion, although their exact character is unknown. (The fourth integral of motion common to spherically-symmetric potentials, an artifact of frequency degeneracy, is destroyed by the perturbation.) It has been long-suspected by astronomers, notably Contopoulos (1967), that the integrals of the motion (apart from the energy) are modified forms of angular momentum. As a paradigm for this problem, Contopoulos considered axisymmetric perturbations to an otherwise spherically-symmetric potential. Then, it is easy to show that the angular momentum component aligned along the symmetry axis is exactly conserved, in addition to the energy. Were it not for the perturbation, the "third integral" would be associated with the total angular momentum (i.e. the magnitude of the angular-momentum vector). Saaf (1968) and Innanen and Papp (1977), among others, have attempted to show that the third integral of motion in aspherical mass distributions is the total angular momentum. Their work, however, did not accommodate fully the nonlinear character of galactic potentials and did not completely resolve the character of stellar orbits and the nature of the third, isolating integral of motion. For this reason, astronomers have employed numerical integrations of the equations of motion to better understand the problem.

Computer-generated stellar orbits do not provide direct information about the integrals of motion. The energy of a particle is conserved and each angular momentum component undergoes complex undulations about some seemingly constant value. Astronomers have adopted two tests for integrability or, conversely, ergodicity. In the first, they plot the stellar orbit projected onto an arbitrary surface, usually a plane in Cartesian geometry which passes through the coordinate origin. On this figure, astronomers also plot a curve that demarcates the region accessible to the particle. For example, this could be a projection of the "zero-velocity surface" (i.e. the surface containing the volume in which energy conservation provides the particle with a positive kinetic energy) onto the plane. If the ergodic hypothesis were to hold, all regions contained within the zero-velocity surface would be accessible to the star. The plots produced by numerical experiments show that the stellar orbits are usually confined to a smaller area on this figure than the ergodic hypothesis would suggest. Contopoulos (1967) has called the characteristic shapes (that demarcate the area filled by stellar orbits) boxes, tubes or shells. These generic forms of orbits can be easily understood in terms of the Stäckel (1891) separable potentials. In the orthogonal coordinates constructed from confocal quadrics, we saw that the motion along each coordinate axis was independent and had a pair of classical turning points. Thus, for example, if we could express the potential in Cartesian coordinates as $U_1(x) + U_2(y) + U_3(z)$ where U_i , $i = 1, 2, 3$ are real-valued functions of the associated coordinate, the stellar orbit would be bounded according to

$$\begin{aligned} x_a &\leq x \leq x_b \\ y_a &\leq y \leq y_b \\ z_a &\leq z \leq z_b \end{aligned} \tag{2}$$

Here, x_a , x_b , etc. are the classical turning points which are related to the isolating integrals (in this geometry, the energies of vibration along each axis) and to the shape of U_1 , U_2 , and U_3 . Equation (2) describes a "box orbit" since the constraints produce a rectangular parallelepiped. In other separable geometries, notably spherical geometry, the pictorial association with a box is somewhat strained. [For spherically symmetric potentials, r , θ , and ϕ have constraints of the form (2), i.e. are individually bounded.] A shell orbit emerges if the constraints in (2), for example, are such that one of the coordinates does not change appreciably. In spherical geometry, confining the radius to a narrow range of values would produce a shell orbit. Tube orbits, on the other hand, result from nearly resonant frequencies associated with the separate coordinates. (Since only ~ 100 orbital periods constitute a galactic lifetime, near-resonance conditions produce tube orbits. If substantially more orbital periods were involved, the tube orbit would assume a box-like character.) These orbits appear very much like a "Lissajous figure." Although this discussion of orbit types is strictly correct only for separable systems, similar behavior is manifest in orbits plotted in projection (in the manner described above) which do not display ergodicity. This first test, then, provides substantial insight into the integrability of the problem.

The second technique, originally suggested by Poincaré, involves a mapping which is most readily applied to two-dimensional problems. It can be applied to galactic problems where the potential is azimuthally symmetric, i.e. does not depend on ϕ (Richstone, 1981). Since the Lagrangian has one ignorable coordinate, the Hamiltonian simplifies to one incorporating two spatial coordinates, the cylindrical coordinates r and z . The mapping, known as the surface of section, is the intersection of the constant energy surface with the equatorial or galactic plane ($z = 0$) and has r and p_r (the momentum conjugate to r) as its coordinates. Each time the star crosses the galactic plane in the upgoing (or, alternatively, the downgoing) direction, a point is made on the (r, p_r) plane. This mapping process is area preserving (a consequence of Liouville's Theorem) and reveals whether or not the motion is integrable. If the system is integrable, the points appear to form a smooth, closed curve. Otherwise, the mapped points explore a two-dimensional region of the (r, p_r) plane. In concluding our discussion of numerical tests for integrability, it is important to distinguish between the mathematical definition of integrability and that adopted by the astrophysical community. To the mathematician, an integral of motion is a quantity that never changes as $t \rightarrow \infty$. To the astrophysicist, an isolating integral is a quantity that has not perceptively changed during the lifetime of the universe (~ 100 stellar orbit periods). In producing projected plots of orbits and surfaces of section, the astrophysicist will cease the numerical integration after 10^{10} years. If by that time no manifestly ergodic behavior is present, he will conclude that the potential is integrable!

The numerical experiments of Schwarzschild (1979) and Richstone (1980, 1981) have provided much insight into galactic dynamics. Schwarzschild's work, in particular, has shown that it is possible to superpose stellar orbits so as to reproduce the underlying mass density distribution. In other words, he obtained a self-consistent (albeit numerical) description of dynamical equilibrium where the density profile corresponded to that of observed elliptical galaxies. Since Schwarzschild's model galaxy had no symmetry axis, surfaces of section were not available as a diagnostic test of integrability. Other numerical tests that he performed, however, provided a strong indication of the existence of a second and third integral of motion. Richstone (1980, 1981) chose a model galaxy possessing a symmetry axis so that the perturbing potential was a function of the galactic latitude (essentially, $\pi/2 - \theta$) and not of ϕ . He was, therefore, able to plot stellar orbits in projection as well as surfaces of section. Of the many cases that he considered, not one was ergodic! In the work of Schwarzschild (1979) and Richstone (1980, 1981), the angular momentum oscillated around a constant value suggesting that the total angular momentum was the leading term in a perturbation expression for the third integral of motion.

The association between angular momentum and integrals of the motion for stellar systems in modest departure from spherical symmetry provides a motivation for exploring the relationship between coordinate geometry and isolating integrals. The standard approach [see, for example, Landau and Lifshitz (1960) or Goldstein (1980)] employs an analysis of the Hamilton-Jacobi equation. We show here that the Hamilton-Lagrange equations of motion can be employed in order to obtain the same integrability conditions, although we shall do so explicitly only for spherical coordinate geometry (the geometry best suited to elliptical galaxies). This method of analysis is especially valuable since dynamical problems are most easily solved using the Lagrange or Hamilton equations of motion and facilitates the use of perturbation methods for ellipsoidal mass distributions.

ISOLATING INTEGRALS AND STRETCHED COORDINATES

Let us begin by considering the motion of a unit mass in the plane under the influence of the potential $V(r, \phi)$. From its Lagrangian

$$L = \frac{1}{2} \dot{r}^2 + \frac{1}{2} r^2 \dot{\phi}^2 - V(r, \phi) \quad , \quad (3)$$

we obtain the momenta p_r and p_ϕ conjugate to r and ϕ ; viz.

$$p_r = \frac{\partial}{\partial \dot{r}} L = \dot{r}, \quad p_\phi = \frac{\partial}{\partial \dot{\phi}} L = r^2 \dot{\phi} \quad ; \quad (4)$$

and the Lagrange equations of motion

$$\ddot{r} = r \dot{\phi}^2 - \frac{\partial}{\partial r} V(r, \phi), \quad \ddot{\phi} = - \frac{\partial}{\partial \phi} V(r, \phi) \quad . \quad (5)$$

[It is assumed that the potential $V(r, \phi)$ and its partial derivatives are everywhere continuous.] The associated Hamiltonian

$$H = \frac{1}{2} p_r^2 + \frac{1}{2} \frac{p_\phi^2}{r^2} + V(r, \phi) \quad (6)$$

is a first integral for the system of equations (4) and (5).

A second integral for the motion can be obtained using Hamilton-Jacobi associated methods. Instead, construct a phase trajectory equation from the equations for ϕ and p_ϕ [the second part of equations (4) and (5)], namely

$$\frac{dp_\phi}{d\phi} = - \frac{r^2}{p_\phi} \frac{\partial V(r, \phi)}{\partial \phi} \quad . \quad (7)$$

In order that a first integral to (7) exist, $V(r, \phi)$ must be expressible as

$$V(r, \phi) = U(r) + r^{-2} w(\phi) \quad (8)$$

(the result we would have obtained had we analyzed the Hamilton-Jacobi equation). Then, the phase trajectory equation (7) becomes

$$\frac{dp_\phi}{d\phi} = - \frac{1}{p_\phi} \frac{dw(\phi)}{d\phi} \quad (9)$$

with the first integral

$$\frac{1}{2} p_\phi^2 + w(\phi) = \frac{1}{2} p_\phi^2 \quad (10)$$

where p_ϕ is a constant. We shall regard p_ϕ as a (conserved) canonical momentum which is conjugate to some angular coordinate that has yet to be identified. In particular, we employ (6), (8) and (10) to express a transformed Hamiltonian

$$H = \frac{1}{2} p_r^2 + \frac{1}{2} \frac{p_\phi^2}{r^2} + U(r) \quad . \quad (11)$$

Two of the Hamilton equations of motion, namely

$$\dot{p}_\phi = 0, \quad \dot{\phi} = \frac{p_\phi}{r^2} \quad (12)$$

confirm the constancy of p_ϕ and isolate the associated coordinate ϕ (not to be confused with ϕ). From the equation (4) for $\dot{\phi}$ together with (10) and (12), we construct the phase trajectory equation

$$\frac{d\phi}{d\phi} = \frac{p_\phi}{p_\phi} = \sqrt{1 - 2w(\phi)/p_\phi^2} \quad . \quad (13)$$

Upon re-examining (10), it should now be clear that the angular momentum p_ϕ oscillates in some fashion [since w is a periodic function of ϕ with a period of 2π] and that p_ϕ is an exactly-conserved quantity with the dimensions of angular momentum. Moreover, p_ϕ represents the energy contained in the angular momentum and the angle-dependent part of the potential. The canonically conjugate coordinate ϕ corresponds to a "stretched" version of the original coordinate ϕ that depends on both the angle-dependent part of the potential and the initial conditions [cf. equation (13)]. By means of this "coordinate stretching" transformation, we have been able to eliminate the angle-dependent part of the potential from the Hamiltonian, a transformation we will now extend to three dimensions.

Consider the motion of a unit mass under the influence of a potential $V(r, \theta, \phi)$. From its Lagrangian

$$L = \frac{1}{2} \dot{r}^2 + \frac{1}{2} r^2 \dot{\theta}^2 + \frac{1}{2} r^2 \sin^2 \theta \dot{\phi}^2 - V(r, \theta, \phi) \quad , \quad (14)$$

we obtain the momenta p_r , p_θ and p_ϕ conjugate to r , θ and ϕ ; namely,

$$\begin{aligned} p_r &= \frac{\partial}{\partial \dot{r}} L = \dot{r} \\ p_\theta &= \frac{\partial}{\partial \dot{\theta}} L = r^2 \dot{\theta} \\ p_\phi &= \frac{\partial}{\partial \dot{\phi}} L = r^2 \sin^2 \theta \dot{\phi} \end{aligned} \quad (15)$$

and the Lagrange equations of motion

$$\begin{aligned} \ddot{p}_r &= r \dot{\theta}^2 + r \sin^2 \theta \dot{\phi}^2 - \frac{\partial V(r, \theta, \phi)}{\partial r} \\ \ddot{p}_\theta &= r^2 \sin \theta \cos \theta \dot{\phi}^2 - \frac{\partial V(r, \theta, \phi)}{\partial \theta} \\ \ddot{p}_\phi &= - \frac{\partial V(r, \theta, \phi)}{\partial \phi} \end{aligned} \quad . \quad (16)$$

The associated Hamiltonian

$$H = \frac{1}{2} p_r^2 + \frac{1}{2} \frac{p_\theta^2}{r^2} + \frac{1}{2} \frac{p_\phi^2}{r^2 \sin^2 \theta} + V(r, \theta, \phi) \quad (17)$$

is an integral of motion for the system of equations (15) and (16).

A second integral of motion emerges from the phase trajectory equation which results from the equations for $\dot{\phi}$ and \dot{p}_ϕ [the last part of equations (15) and (16)], namely

$$\frac{dp_\phi}{d\phi} = - \frac{r^2 \sin^2 \theta}{p_\phi} \frac{\partial V(r, \theta, \phi)}{\partial \phi} \quad . \quad (18)$$

In order that an integral to (18) exist, $V(r, \theta, \phi)$ must be expressible as

$$V(r, \theta, \phi) = W(r, \theta) + \frac{1}{r^2 \sin^2 \theta} w(\phi) \quad . \quad (19)$$

The phase trajectory equation and its integral are the same as those encountered before, i.e. equations (9) and (10). Defining p_ϕ and ϕ using equations (10), (12) and (13), the Hamiltonian (17) becomes [employing (19)]

$$H = \frac{1}{2} p_r^2 + \frac{1}{2} \frac{p_\theta^2}{r^2} + \frac{1}{2} \frac{p_\phi^2}{r^2 \sin^2 \theta} + W(r, \theta) \quad . \quad (20)$$

The coordinate stretching applied in defining ϕ and p_ϕ have rendered the Hamiltonian cyclic in the variable ϕ .

A third integral of motion emerges from the phase trajectory equation which results from combining the equations for $\dot{\theta}$ and \dot{p}_θ [the middle part of equations (15) and (16)] with equation (19). In particular, we find

$$\frac{dp_\theta}{d\theta} = \frac{r^4 \sin \theta \cos \theta \dot{\phi}^2 - r^2 \frac{\partial W(r, \theta)}{\partial \theta} - \frac{\partial}{\partial \theta} \left(\frac{w(\phi)}{\sin^2 \theta} \right)}{p_\phi} \quad (21)$$

which can be expressed using (10) and (15) as

$$\frac{dp_\theta}{d\theta} = - \frac{r^2 \frac{\partial W(r, \theta)}{\partial \theta} + \frac{\partial}{\partial \theta} \left(\frac{1}{2} \frac{p_\phi^2}{\sin^2 \theta} \right)}{p_\theta} \quad (22)$$

Since p_ϕ is a constant of the motion, a necessary and sufficient condition for the existence of a third integral of the motion is that $W(r, \theta)$ be expressible as

$$W(r, \theta) = U(r) + r^{-2} v(\theta) \quad . \quad (23)$$

Combining this result with (8), we see that

$$V(r, \theta, \phi) = U(r) + \frac{1}{r^2} v(\theta) + \frac{1}{r^2 \sin^2 \theta} w(\phi) \quad , \quad (24)$$

the integrability condition we would have obtained had we solved the Hamilton-Jacobi equation. Thus, the phase trajectory equation (22) now becomes

$$\frac{dp_\theta}{d\theta} = - \frac{1}{p_\theta} \frac{d}{d\theta} \left\{ v(\theta) + \frac{1}{2} \frac{p_\phi^2}{\sin^2 \theta} \right\} \quad (25)$$

with the integral

$$\frac{1}{2} p_\theta^2 + \frac{1}{2} \frac{p_\phi^2}{\sin^2 \theta} + v(\theta) = \frac{1}{2} p_\theta^2 \quad (26)$$

where p_θ is a constant. Note that $v(\theta) + \frac{1}{2} \frac{p_\phi^2}{\sin^2 \theta}$ isolates the latitude-dependent part of the potential, which we shall call $\tilde{v}(\theta)$, and introduces latitudinal

turning points into the problem if p_ϕ (a constant of the motion associated with the azimuthal angular momentum) is non-zero. As before, we shall regard p_θ as a (conserved) canonical momentum which is conjugate to some angular coordinate θ (not to be confused with ϕ) that has yet to be identified. We employ equations (20), (23) and (26) to express a transformed Hamiltonian

$$H = \frac{1}{2} p_r^2 + \frac{1}{2} \frac{p_\theta^2}{r^2} + U(r) \quad . \quad (27)$$

From this Hamiltonian, we associate p_θ with the total angular momentum (and a "centripetal barrier") just as we associate p_ϕ [via (26)] with the azimuthal component of the angular momentum. Two (new) Hamilton equations of motion describe p_θ and θ , namely

$$\dot{p}_\theta = 0, \quad \dot{\theta} = \frac{p_\theta}{r^2} \quad . \quad (28)$$

This again confirms the constancy of p_θ and provides a definition for $\dot{\theta}$. From equation (15) for $\dot{\phi}$ together with (26) and (28), we construct the phase trajectory equation

$$\begin{aligned} \frac{d\theta}{d\phi} &= \frac{p_\theta}{p_\phi} = \frac{\sqrt{1 - 2\tilde{v}(\theta)/p_\theta^2}}{\theta} \\ &= \sqrt{1 - [2v(\theta) + p_\phi^2]/p_\theta^2} \end{aligned} \quad (29)$$

which provides a more useful definition of θ . From equation (26), it follows that the (true) total angular momentum p_θ oscillates whereas p_ϕ , which is an exactly conserved quantity with the dimensions of angular momentum, does not. For the three-dimensional problem, p_θ represents the energy contained in all of the angular momentum components and the angle-dependent part of the potential. The new stretched coordinate θ has rendered the Hamiltonian (27) cyclic in all angular coordinates.

The three integrals that emerge in spherical geometry are the energy [or Hamiltonian (27)] and quantities related to the total angular momentum

$$p_\theta = r^2 \dot{\theta} \quad (28')$$

and its azimuthal component

$$p_\phi = r^2 \sin^2 \theta \dot{\phi} \quad (12')$$

[the latter includes the latitudinal dependence of three-dimensional problems not given in (12)]. Equations (10) and (26) show that we can have independent turning points in our two angular coordinates. For realistic galactic potentials, the radial potential $U(r)$ will cause the Hamiltonian (27) to produce two radial turning points (a centripetal and potential barrier, respectively). Out of these, the box, shell and tube orbits (the latter currently referred to a "pipes") described earlier arise naturally.

The phase trajectory methods just employed can also be adapted to other separable potentials of the Stäckel type, although we do not do so here. In this communication, we have surveyed a few noteworthy developments in stellar dynamics and have exploited the relation between dynamics and the geometry of curved spaces. The most exhaustive investigation of this relationship was due to Synge (1926) who employed methods of differential geometry and tensor calculus. Indeed, there are grounds for cautious optimism that the further application of these techniques together with the topologically-oriented methods of the "new nonlinear dynamics" will yield a substantial increase in our understanding of galactic dynamics and

structure.

This work was initiated while at the Institute for Advanced Study, Princeton, N.J. In preparing this paper, the author has benefitted greatly from conversations with F. Dyson, J.N. Bahcall, R.G. Newton, S.D. Tremaine, M. Schwarzschild, D.O. Richstone, W.M. Kaula and F.H. Busse. This research was supported by a grant from the California Space Institute to the Department of Earth and Space Sciences, University of California.

REFERENCES

- [1] Berry, M.V., Regular and Irregular Motion, in: S. Jorna (ed.), Topics in Nonlinear Dynamics, American Institute of Physics Conference Proceedings, No. 46 (1978).
- [2] Contopoulos, G., Problems of Stellar Dynamics, in: J. Ehlers (ed.), Relativity and Astrophysics, 2. Galactic Structure (American Mathematical Society, Providence, R.I., 1967).
- [3] Eisenhart, L.P., Separable Systems of Stäckel, Ann. Math. 35 (1934) 284-305.
- [4] Goldstein, H., Classical Mechanics (Addison-Wesley, Reading, Mass., 1980).
- [5] Innanen, K.A. and Papp, K.A., Particle Dynamics in Spheroidal Mass Distributions, the Third Integral of Motion and Angular Momentum, Astronom. J. 82 (1977) 322-328.
- [6] Kolmogorov, A.N., The General Theory of Dynamical Systems and Classical Mechanics, text of address to the 1954 International Congress of Mathematicians.
- [7] Landau, L.D. and Lifshitz, E.M., Mechanics (Oxford: Pergamon, 1960).
- [8] Lynden-Bell, D., Stellar Dynamics Potentials with Isolating Integrals, Monthly Notices Roy. Astronom. Soc. 124 (1962) 95-123.
- [9] Morse, P.M. and Feshbach, H., Methods of Theoretical Physics (McGraw-Hill, New York, 1953).
- [10] Pars, L.A., A Treatise on Analytical Dynamics (Wiley, New York, 1965).
- [11] Richstone, D.O., Scale-Free, Axisymmetric Galaxy Models with Little Angular Momentum, Astrophys. J. 238 (1980) 103-109.
- [12] Richstone, D.O., Scale-Free Models of Galaxies II: A Complete Survey of Orbits, Department of Astronomy preprint, University of Michigan.
- [13] Saaf, A.F., A Formal Third Integral of Motion in a Nearly Spherical Stellar System, Astrophys. J. 154 (1968) 483-498.
- [14] Stäckel, P.G., Ueber die Integration des Hamilton-Jacobischen Differentialgleichungen Mittelst separation der variablen, Habilitationsschrift, Halle (1891).
- [15] Schwarzschild, M., A Numerical Model for a Triaxial Stellar System in Dynamical Equilibrium, Astrophys. J. 232 (1979) 236-247.
- [16] Synge, J.L., On the Geometry of Dynamics, Philos. Transactions (A) 226 (1927) 31-106.

This Page Intentionally Left Blank

PART II
Nonlinearity in Field Theories
and Low Dimensional Solids

This Page Intentionally Left Blank

PHYSICS IN FEW DIMENSIONS

V. J. Emery

Physics Department *
Brookhaven National Laboratory
Upton, New York 11973

This article is a qualitative account of some aspects of physics in few dimensions, and its relationship to nonlinear field theories. After a survey of materials and some of the models that have been used to describe them, the various methods of solution are compared and contrasted. The roles of exact results, operator representations and the renormalization group transformation are described, and a uniform picture of the behavior of low-dimensional systems is presented.

I. INTRODUCTION

Many of the fundamental problems of condensed matter physics may be regarded as examples of nonlinear field theories. This point of view has been advantageous for the quantum-mechanical many-body problem and for the modern theory of phase transitions, leading to successful theories of single-particle and collective phenomena in a wide variety of physical systems. It is the basis for our current understanding of superconductivity and superfluidity, and has led to a deeper understanding of the universal cooperative effects which are observed in the neighborhood of a critical point. Most of these developments have relied upon the techniques of quantum field theory, and have made little reference to concepts that have arisen in a purely classical context. More recently, however, this situation has changed, and some of the approaches to physics in one and two space dimensions have been much closer in spirit to the ideas of classical nonlinear field theory; references to the sine-Gordon equation, solitons, breathers, inverse scattering, etc., have become relatively commonplace in the literature of elementary particle and condensed matter physics.

A feeling for the relevance of these concepts and their supplementary relationship to the more "conventional" approach may be obtained by considering several examples from condensed matter physics. In what follows, there will be no reference to the theory of non-integrable systems, although its influence is beginning to make itself felt, for example, in attempts to understand some aspects of the theory of incommensurate structures¹.

The systems of interest are quite diverse in their physical characteristics. Some are quantum mechanical and one dimensional, others are classical and two dimensional. But it turns out that there is a common thread to their mathematical formulation: a remarkably large number of models are equivalent to each other, either exactly or in the asymptotic properties which govern the kinds of long-ranged order that might be established. All bear some relationship to the sine-Gordon equation.

The impetus for these developments has come from both theory and experiment, from the desire for a unified view of two-dimensional models as well

as a need to understand a number of unusual observations. The ingenuity of the synthetic chemist and the experimental physicist has been a continual driving force in the whole field, and it is appropriate to start out with some brief mention of the materials they have investigated, before going on to describe the mathematical models, and the methods that have been devised in order to solve them.

II. MATERIALS

Low-dimensional behavior arises in two principal ways. Some materials are extremely anisotropic, consisting of atoms, molecules or ions arranged in chains or in layers, which are weakly coupled to their environment, and act independently over a wide range of temperatures. The structure imposes its own constraints on the motion of electrons, and this leads to characteristic low-dimensional behavior of the electrical properties. The alternative is to have a restrictive geometry; a free film or a film adsorbed on a surface are essentially two dimensional, whereas particles confined to narrow channels may have a one dimensional character. In most cases there are circumstances in which the true three-dimensional nature of the system makes itself felt, and understanding the crossover to this regime is part of the interest in the problem. The purpose of studying these systems is to look for phenomena that may not occur, or may be difficult to produce in more isotropic materials: effects of disorder are expected to be more pronounced, and certain kinds of phase transition may take place more readily. It has long been a hope to find superconductivity at relatively high temperatures in anisotropic organic materials, where strong electron-electron interactions may be produced by an excitonic mechanism².

All of these effects are likely to involve an interplay between the many degrees of freedom which reside in a molecular crystal--the spin and translation of electrons or the spin, orientation, vibration and translation of the molecules. For this reason, it is often quite difficult to extract, from rather indirect experimental information, the primary mechanism which drives the behavior of a particular physical system. The common approach has been to solve a number of simplified models, in order to discover the particular effects to look out for and to limit the possible range of explanations of a given experiment.

A few examples will illustrate the nature of the systems which have been investigated. One-dimensional materials frequently contain rather flat organic molecules³ such as TTF (tetrathiafulvalene), TSeF (tetraselenofulvalene), TCNQ (tetracyanoquinodimethane) and TMTSF (tetramethyltetraselenofulvalene), all of which may be arranged in closely packed stacks. In TTFCuBDT, long-ranged correlations in chains of localized spins conspire with peculiarities of the lattice vibrations to produce a dimerized state⁴ (spin-Peierls transition). The organic metals TTF-TCNQ, TSeF-TCNQ and (TMTSF)₂PF₆ are electrical conductors because charge transfer from donor to acceptor molecules leaves a partially filled band of states³. The conductivity of these systems increases to a quite high value as the temperature is lowered, but ultimately this is reversed by a metal-insulator transition. However, at sufficiently high pressure, (TMTSF)₂PF₆ becomes a superconductor. An example of a so-called molecular metal is Hg_{3-δ}AsF₆, which consists of chains of Hg ions arranged in planes and interspersed with AsF₆⁻ ions. Its peculiar properties are a consequence of the lack of commensurability between the Hg chains and the AsF₆⁻ lattice (δ is a function of temperature and is about 0.2). At room temperature, the weakly coupled Hg chains form one-dimensional liquids⁵, but they freeze at about 120 K. This transition is unusual in that it is an example of continuous freezing. At much lower temperatures the material becomes a superconductor⁶, but the origin and properties of this state are not fully understood. An entire session of this conference is concerned with the properties of polyacetylene, a linear system which is thought to form a dimerized chain with soliton dislocations.

Much of the recent interest in two-dimensional materials has been centered upon a class of systems which have a phase transition but are unable to establish the related long-ranged order because of the destructive effect of thermal fluctuations. They accomplish this as the temperature is decreased below a certain value, by remaining on the verge of a transition to an ordered state: every point is critical. Solids, superconductors, superfluids and some phases of liquid crystals are expected to have this behavior in two dimensions, and verification has been sought in adsorbed layers or freely-suspended films. Another interesting property of overlayers is the existence of periodic structures that may be commensurate or incommensurate with the substrate lattice. This problem has been investigated by the scattering of neutrons, x-rays and electrons as well as by photoemission experiments, and, although a picture of what is going on has steadily been developed, it is still far from complete. Other two-dimensional forms of spin ordering, displacive phase transitions and charge-density wave states are to be found in bulk materials with a layered structure. A more extensive review of this whole subject may be found in the proceedings of the 1979 Kyoto summer school⁷ and the 1980 Lake Geneva, Wisconsin conference⁸.

III. MODELS

At first sight it seems inappropriate to pay so much attention to crystalline solids and lattice models in a discussion of nonlinear field theories, but there is much to be gained from relating one to the other by taking the continuum limit.

The location of a point in a simple cubic lattice is specified by a vector $\vec{r} = s\vec{n}$, where \vec{n} has integer components and s is the lattice spacing. Correlation functions have a characteristic length scale $s\xi$, where ξ , the coherence length of the lattice model, is a pure number. A field theory is obtained by taking the continuum limit $s \rightarrow 0$, in which finite differences become derivatives, and lattice sums become integrals. The advantage of considering this limit, which is clearly fictitious for a real solid, is that it focuses attention on the asymptotic properties of correlation functions, and forces us to consider behavior at a critical point. For if the length scale $s\xi$, and position vector \vec{r} are to remain finite as $s \rightarrow 0$, it is necessary that $\xi \rightarrow \infty$ (which is the case at a critical point) and $|\vec{n}| \rightarrow \infty$ (which is the asymptotic limit). Equally, the continuum limit is useful for a field theory when the critical properties of the corresponding lattice model are known.

In the present context, however, there is a further important reason for taking the continuum limit--it is crucial for establishing the relationship between models and obtaining a unified picture of physics in two dimensions. Space does not permit an account of the technical developments necessary for the implementation of this program, but a survey of some of the interesting Hamiltonians should give at least some idea of the kinds of system under consideration. Classical, two-dimensional models will be described first. Potts Models⁹

$$H_p = -J \sum_{n,n.} \delta_{S_i S_j} \quad (1)$$

with $S_i = 1, 2, \dots, q$. Here, the summation is carried over near-neighbor sites i, j on a square lattice. The state of lowest energy has all spins equal and is q -fold degenerate. This Hamiltonian is relevant for adsorbed films⁹ and for magnetic systems. ($q=2$ is identical to the Ising model)

Ashkin-Teller Model¹⁰

$$H_A = - \sum_{nn} \{J_1 S_i S_j + J_2 T_i T_j + J_3 S_i T_i S_j T_j\} \quad (2)$$

where $S_i = \pm 1$, $T_i = \pm 1$. This is a two-component lattice gas and, for $J_1 = J_2 = 0$, it is identical to the 4-state Potts model.

Interface Roughening¹¹

$$H_R = - \sum_{n,n} f(h_i - h_j) \quad (3)$$

with $h_i = 0, \pm 1, \pm 2, \dots$, is a cell model of crystal growth. The constituents are assumed to fill a column of cells to a height h_i at lattice site i . For the physical model, $f(h) = h$ (solid-on-solid model), but another case of particular importance is $f(h) = h^2$ (discrete Gaussian model) which is directly related to the Coulomb gas and the xy -model, as will be seen.

Vertex Models¹²

It is imagined that every vertex in a square lattice is connected to its four neighbors by a link, which has a sense (right or left, up or down) specified by an arrow. There are sixteen different kinds of vertex (four links, each with two senses) and each is assigned a different weight. The problem is to sum over all configurations of links. The eight-vertex problem (in which each vertex has an even number of incoming and outgoing arrows) has been solved exactly by Baxter¹². It is also of interest to consider a more general staggered version of this model with two sets of weights, one for each of two interpenetrating sublattices. Such a model has been shown to be equivalent to the Potts models¹³ and to the Ashkin-Teller model¹⁰.

The xy -Model¹⁴

which has a Hamiltonian given by

$$H_{xy} = -J \sum_{n,n} \vec{V}_i \cdot \vec{V}_j \quad (4)$$

where \vec{V}_i is a unit, two-dimensional vector, differs from the preceding models by having a continuous variable at every site. This is clear if $\vec{V}_i \cdot \vec{V}_j$ is rewritten in the form $\cos(\theta_i - \theta_j)$ where θ_i and θ_j are the polar angles of \vec{V}_i and \vec{V}_j . This Hamiltonian describes a magnetic system, but it has also been used for the superfluid transition in He^4 films, for which θ_i is the phase of the order parameter. The dual¹⁴ of H_{xy} is a special case of H_R .

Coulomb Gas¹⁴

$$H_C = - \sum_{\text{all } i,j} \sum_{Q_i} Q_i Q_j \ln |\vec{r}_i - \vec{r}_j| \quad (5)$$

Here the summation over i and j extends to all sites (not only near neighbors), and the charges Q_i have values $0, \pm 1, \pm 2, \dots$. This model is equivalent^{14,16} to the discrete Gaussian version of H_R , as mentioned earlier.

It is a remarkable fact that these apparently quite different models are closely related: the partition function of one may be transformed into the partition function of the other, with an appropriate redefinition of parameters. In some cases, the transformation is exact in others it is asymptotically correct for the critical properties⁹⁻¹⁶. It is often possible to find further equivalences between correlation functions. However, the transition to field theory is most directly made via the transformation of all of these models into one-dimensional quantum mechanical systems. The link is provided by the transfer matrix¹⁷ T .

The partition function Z is a sum over all configurations of variables on the lattice. The elements of T consist of the contributions to this sum from pairs of configurations of two neighboring rows of the lattice, and

$$Z = \text{Tr } T^N \quad (6)$$

where N is the number of rows. In the thermodynamic limit ($N \rightarrow \infty$), Z is dominated by the largest eigenvalue of T . The configurations of a row may also be regarded as states of a one-dimensional quantum system, with T playing the role of a transition matrix. In this interpretation, if T is written in the form¹⁸ $\exp(-H)$, then H is the corresponding quantum Hamiltonian, and its ground state gives the largest eigenvalue of T . It may appear that one of the original space dimensions has been lost, but actually it has been replaced by (imaginary) time which, implicitly or explicitly, plays an unavoidable role in quantum mechanics. Following this procedure, every one of the models listed above may be related to the spin Hamiltonian¹⁹

$$H = H_0 + H_1 + H_2 \quad (7)$$

where

$$H_0 = - \sum_{j=1}^N [\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y - g \sigma_j^z \sigma_{j+1}^z] \quad (8)$$

$$H_1 = \gamma \sum_{j=1}^N [\sigma_j^x \sigma_{j+1}^x - \sigma_j^y \sigma_{j+1}^y] \quad (9)$$

and

$$H_2 = -\lambda \sum_{j=1}^N (-1)^j [\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y - g' \sigma_j^z \sigma_{j+1}^z + h_s \sigma_j^z] \quad (10)$$

Here σ_j^x , σ_j^y and σ_j^z are Pauli matrices. The Hamiltonian H_0 describes the Heisenberg-Ising model, H_1 gives an anisotropy in the xy plane of spin space and H_2 is a dimerization that is related to the staggering of weights in the Baxter model. This is the central model of the field to which all others may be reduced, at least asymptotically¹⁹. The spin representation (8)-(10) is only one way of writing H . Other useful forms in terms of fermion or boson variables will be introduced later. The parameters, g , γ , λ , g' , h_s are known functions of the temperature and the parameters (J , q , etc.) of the original models. It will be seen that H_0 is the critical Hamiltonian while H_1 and H_2 give thermal or field perturbations away from the critical points of the original models.

It is now possible to state how the quantum models, mentioned in Section 2, may be fitted into the picture. The Hamiltonian for a spin-Peierls system⁴ is $H_0 + H_2$, where H_0 refers to localized spins in a uniform lattice, and H_2 describes the effects of dimerization in the low-temperature phase. The motion of the mercury ions in $\text{Hg}_3\text{-gAsF}_6$, and the spin or charge degrees of freedom of electrons in organic conductors are related to $H_0 + H_1$, but, to show more

explicitly how this comes about, it is necessary to know something about transformations between spins, bosons and fermions. This is described in the next section.

IV. METHODS OF SOLUTION

It is a remarkable feature of one-dimensional physics that, for a number of models, eigenstates and eigenvalues are known exactly. They have been obtained in various equivalent forms--Bethe's ansatz for the wavefunction (see Dr. Andrei's talk), quantum inverse scattering²⁰ and the semiclassical method for field theory²¹. This whole approach is closely related to the ideas of classical nonlinear physics, and it works for systems which are exactly integrable.

Once an exact solution is available, it might seem that there is little more to be said. However it is not easy to work with the wavefunctions and, with one exception²², it has not been possible to evaluate correlation functions which are required in order to assess the prospects for various kinds of long-ranged order. Furthermore, the method has not so far succeeded for the most general Hamiltonian of Eqs. (6)-(10), including dimerization, and hence there is every reason to seek alternative approaches, even approximate ones, that are not so specific. Two are of particular importance--operator representations, and the renormalization group method.

The idea of using operator representations is that there are exact relationships between spin, fermion and boson operators, and it may happen that a problem is intractable in one representation but may be exactly soluble in another. Perhaps the best-known example is the Jordan-Wigner transformation²³

$$\sigma_m^+ = \exp(i\pi \sum_{j=1}^{m-1} c_j^\dagger c_j) c_m^\dagger \quad (11)$$

$$\sigma_m^- = \exp(i\pi \sum_{j=1}^{m-1} c_j^\dagger c_j) c_m \quad (12)$$

where

$$\sigma_m^\pm = 1/2(\sigma_m^x \pm i\sigma_m^y) \quad (13)$$

and c_m^\dagger, c_m are fermion creation and annihilation operators. This transformation is used in solving the two-dimensional Ising model²³, and it may be used for any lattice model. On the other hand the boson representations of spin or fermion operators²⁴ rely on the continuum limit²⁵. For fermions in one dimension, it is possible to distinguish between right-going and left-going particles, with field operators $\psi_+(x)$ and $\psi_-(x)$ respectively. Then $\psi_\pm(x)$ may be written²⁴

$$\psi_\pm(x) = \text{const} \exp[-i(\pi/\mu)^{1/2} \int_{-\infty}^x d\xi \pi(\xi) \pm i(\pi\mu)^{1/2} \phi(x)] \quad (14)$$

where $\phi(x)$ is a Bose Field and $\pi(x)$ is its conjugate momentum. This transformation is particularly useful, because it is rather easy to evaluate correlation functions when the operators consist of exponentials of Bose Fields²⁴. Equations (11)-(14) are given in order to show the form of the transformations. More detailed discussions and applications to a number of problems may be found in the literature or in reviews^{23,24,25}.

The operator representations give the connection between the spin chain and other one dimensional materials, and also show how the sine-Gordon equation comes into the picture. In the continuum limit, the charge- and spin-degrees of freedom of the conduction electrons are decoupled²⁴, and each may be regarded as a set of spinless fermions which, in turn, are related to $H_0 + H_1$ by a Jordan-Wigner transformation. The Hamiltonian contains products of $\psi_{\pm}(x)$ and $\psi_{\pm}^{\dagger}(x)$ and, introducing the boson representation (14), it is found that the contribution from the integral of $\pi(\xi)$ cancels out leaving a factor proportional to $\cos[(\pi\mu)^{1/2}\phi(x)]$, which is the potential energy density of the sine-Gordon system. Thus, all of these problems, as well as the ordered phase of mercury ions in $\text{Hg}_{3-8}\text{AsF}_6$ (another sine-Gordon system²⁵) are related to the spin chain.

The sine-Gordon equivalence suggests that there may be soliton and breather excitations. However, the fields are quantized and, in contrast to the classical case^{21,26}, the solution depends on the value of μ . Solitons and breathers exist when μ is less than a critical value, and then the mass of the lowest excitation is a measure of the coherence length in the disordered phase of the corresponding two dimensional problem; since the mass governs the decay in (imaginary) time of a quantum-mechanical system. It is in this region that the sine-Gordon picture contributes most effectively to the solution of this group of problems²⁴. If μ is greater than the critical value, there are no solitons and the excitations are massless. This region corresponds to the line of critical points in the two dimensional theories²⁷, and is more effectively tackled by the renormalization group method.

One way of phrasing the renormalization group method²⁸ is to focus on some quantity such as the coherence length ξ , and to study the variation of parameters (such as J) required to keep ξ s fixed, as s varies. From the resulting flow equations, it is possible to evaluate the critical exponents. Usually this procedure cannot be carried out exactly unless there is a small parameter in which an expansion may be made. It does, however, tell us which are the relevant variables, the ones which must be taken into account in order to get a complete description. The method is most useful in the neighborhood of a fixed point of the transformation, and it is usually necessary to resort to numerical calculations or to some other method of calculation if a more global picture is required. Nevertheless it can be applied in a relatively straightforward way to more complex Hamiltonians, and to include the dimerization H_2 , which is quite difficult to deal with otherwise¹⁹.

Clearly, all of these approaches have their advantages and limitations, and it is necessary to resort to a combination of all of them, in order to build up a complete picture of a given problem. The Bethe ansatz or the quantum versions of the inverse scattering method are directly related to classical nonlinear theories. They give exact expressions for the energy spectrum, but it is quite difficult to evaluate correlation functions. The transformations between spins, fermions and bosons may help to turn a problem into a more easily soluble form but, without further help, they do not always give a mass spectrum. They are at their best near a critical point, where the boson form in particular is useful in giving the algebraic decay of correlation functions and expressions for the associated critical exponents²⁴. The renormalization group was originally a technique of quantum field theory. It is versatile but does not give a complete analytical solution to a problem if there is no small parameter in which to expand, or if regions far from a fixed point cannot be disregarded.

Nevertheless, by assembling the contributions of all of these techniques, we have come to a unified picture of a large class of one-or two-dimensional

models. The common feature is a line of critical points, along which correlation functions of the classical or quantum models decay algebraically; with exponents that depend upon the position along the line. The boson representations and renormalization group equations lead to relationships between critical exponents in what amounts to an extension of the exact results to problems for which no exact solution exists¹⁹. Off the critical line, it is necessary to know the mass spectrum or the coherence length and to make use of all of the methods in order to obtain a solution. A survey of this approach and a description of recent work is given in reference 19.

All of these developments have been described in the context of condensed matter physics, but many of the results have been discovered independently by elementary particle theorists. Their objective has been to practice on models showing confinement and asymptotic freedom in the hope that techniques may be of value for the more physical four dimensional theories. By now it has the appearance of a mature field. But it is too much to expect that such a tidy picture of low-dimensional physics will persist. Already a number of models, that cannot be solved immediately by these methods, are being investigated--that is a symptom of a healthy field.

REFERENCES

* Research supported by DOE under Contract No.DE-AC02-76CH00016.

- [1] Aubry, S., in Solitons and Condensed Matter Physics, Edited by Bishop A. R. and Schneider, T. (Springer-Verlag, Berlin, Heidelberg, New York, 1978).
- [2] Little, W. A., Phys. Rev. 134A (1964) 1416-1424.
- [3] See for example articles in Chemistry and Physics of One Dimensional Metals, Edited by Keller, H. J. (Plenum, New York, 1977); Molecular Metals, Edited by Hatfield, W. E. (Plenum, New York, 1979).
- [4] Cross, M. C. and Fisher, D. S., Phys. Rev. B19 (1979) 402.
- [5] Emery, V. J. and Axe, J. D., Phys. Rev. Lett. 40 (1978) 1507.
- [6] For a review see Heeger, A. J. and MacDiarmid, A. G. in Molecular Metals, reference 3, p. 419.
- [7] Proceedings of the Kyoto Summer Institute 1979-Physics of Low-Dimensional Systems, edited by Nagaoka, Y. and Hikami, S. (Publication Office, Progress of Theoretical Physics, 1979).
- [8] Ordering in Two Dimensions, Edited by Sinha, S. K. (North Holland, New York, Amsterdam, Oxford, 1980).
- [9] Potts, R. P., Proc. Camb. Phil. Soc. 48 (1952) 106.
- [10] Wegner, F. J., J. Phys. C5 (1972) L131.
- [11] See e.g. Emery, V. J. and Swendsen, R. H., Phys. Rev. Lett. 39 (1977) 1414.
- [12] For a review see M. P. M. den Nijs, J. Phys. A12 (1979) 1857.
- [13] Temperley, H. N. V. and Lieb, E. H., Proc. Roy. Soc. London Ser. A322 (1971) 251.
- [14] Jose, J. V., Kadanoff, L. P., Kirkpatrick, S. and Nelson, D. R., Phys. Rev. B16 (1977) 1217.
- [15] van Beijeren, H., Phys. Rev. Lett. 38 (1977) 993.
- [16] Chui, S. T. and Weeks, J. D., Phys. Rev. B14 (1976) 4978.
- [17] This is a very common method of solving problems in two-dimensional statistical mechanics, specific examples are given in references 13 and 23.
- [18] Formally this is correct in the limit of strong coupling in one direction and weak coupling in the other. It is often used as an approximate procedure to find a Hamiltonian with the same critical properties as the classical model. A more detailed discussion is given in reference 19.

- [19] Black, J. L. and Emery, V. J., Phys. Rev. B23 (1981) 429 and article to be published.
- [20] For a review see Fowler, M. in Physics in One Dimension, Edited by J. Bernasconi and Schneider, T., Springer Series in Solid-State Sciences (Springer, Berlin, Heidelberg, New York, 1980).
- [21] Dashen, R. F., Hasslacher, B. H. and Neveu, A., Phys. Rev. D11 (1975) 3424.
- [22] Johnson, J. D., Krinsky, S. and McCoy, B. M., Phys. Rev. A8 (1972) 2526.
- [23] Schultz, T. D., Mattis, D. C. and Lieb, E. H., Rev. Mod. Phys. 36 (1964) 856.
- [24] Emery, V. J. in Highly Conducting One-Dimensional Solids, p. 247, Edited by de Vreese, J. T., Evvard, R. P. and van Doren, V. E. (Plenum, New York 1979).
- [25] Emery, V. J. in Proceedings of the Kyoto Summer Institute 1979, p. 1 (reference 7).
- [26] Luther, A., Phys. Rev. B15 (1977) 403.
- [27] Moving along the line of critical points may correspond to changing the temperature, or parameters of the model such as vertex weights or the number (q) of states of the Potts model. Thus the various models may have different exponents because they are related to different points on the critical line.
- [28] See for example Introduction to the Renormalization Group and Critical Phenomena, Pfeuty, P. and Toulouse, G. (Wiley, New York 1977).

This Page Intentionally Left Blank

SOLUTION OF THE KONDO PROBLEM

N. Andrei

Department of Physics, New York University
New York, New York 10003 USA

and

Department of Physics & Astronomy, Rutgers University
Piscataway, New Jersey 08854 USA

The Kondo Hamiltonian is shown to be exactly soluble. We construct the spectrum, calculate the magnetization curve and present an analytical determination of Wilson's number.

In this talk I shall present an exact solution for the Kondo Hamiltonian

$$= -i \int \phi_a^* \partial_x \phi_a(x) dx + J \vec{S} \phi_a^*(0) \vec{\sigma}_{ab} \phi_b(0) \quad , \quad J > 0 \quad (1a)$$

Here $\phi_a(x)$ is an electron field with spin components $a = \pm \frac{1}{2}$ and \vec{S} is a spin operator.

The Hamiltonian represents a gas of electrons interacting via spin exchange with an impurity localized at $x=0$, and described by the spin operator \vec{S} .

It has been extensively studied in the past [1] and many of its properties were understood. Let me summarize those aspects that are relevant to our problem.

To start with, we shall concentrate only on quantities that are independent of the cut-off D . In other words, as is the case with every renormalizable field theory, requires a cut-off in order to be properly defined. There are various ways of introducing cut-offs and as long as they are kept finite, there are properties that do depend on them and there are properties that do not. We shall investigate only the latter which are called universal quantities.

To effect this we shall restrict the temperature T and the magnetic field H to be small compared with the cut-off D , ($T, H \ll D$). This will be termed the universal or scaling regime. It still turns out [2] that the model does have a scale $T_0 = T_0(D, J)$ (to be defined later) which is relevant in the scaling regime, and which uniquely characterizes the system. In other words, if $F = F(T, H; D, J)$ is the free energy, then in the scaling region

$$\frac{F}{T}(T, H; D, J) \rightarrow f\left(\frac{T}{T_0}, \frac{H}{T_0}\right)$$

where f is universal, and the only place where the details of the construction enter is the form of T_0 and its dependence on the cut-off and coupling constant, which is, usually, of the form $T_0 = D e^{J^{-1}}$. To keep T_0 invariant the coupling constant must depend on the cut-off in such a way that it becomes logarithmically smaller with an increasing cut-off, $J = (\log \frac{D}{D_0})^{-1}$ and the manner it approaches zero depends on the choice of T_0 . This property is called asymptotic freedom[3][4].

In asymptotically free theories, however, one may apply perturbation theory at high temperatures or magnetic fields $T, H \gg T_0$ (but still $T, H \ll D$). Thus the high temperature regime is accessible by conventional methods. But when one crosses over to low temperatures ($T \ll T_0$) one reaches a strong coupling regime [2] and the approximation schemes break down. To quantify this, consider the impurity susceptibility χ^i . For a free spin $\chi^i = \frac{\partial M^i}{\partial H} \Big|_{H=0} = \frac{\mu^2}{T}$ (Curie's Law) and one expects to find, even when interactions are present, that the susceptibility approaches this value. Thus

$$\chi^i \xrightarrow{T \gg T_0} \frac{\mu^2}{T} \left[1 - \frac{1}{\log \frac{T}{T_K}} - \frac{1}{2} \frac{\log \log \frac{T}{T_K}}{\log^2 \frac{T}{T_K}} + O\left(\frac{1}{\log \frac{T}{T_K}}\right)^3 \right]$$

where we see the logarithmic corrections typical to asymptotic freedom. We have also introduced a scale T_K which is defined by absorbing the $(\log \frac{T}{T_0})^{-2}$ term into the $(\log \frac{T}{T_0})^{-1}$ term.

Crossing over to zero temperature one finds [2] that the susceptibility, rather than diverging like $\frac{\mu^2}{T}$, is finite and defines a new scale T_0

$$\chi^i \Big|_{T=0} = \frac{\mu^2}{\pi T_0} \quad .$$

The physical mechanism that renders χ^i finite is screening (quenching) which leads to a spinless impurity and hence to a finite χ^i .

The ratio $W = T_K/T_0$ is a pure calculable number [4]. It characterizes the cross-over in the properties of a system from a weak to a strong coupling regime (similar to QCD), and requires non-perturbative calculation. It was first found numerically by Wilson [1] and we shall derive [5] an analytic expression, using an exact diagonalization [6] [7] of the Hamiltonian to which we now turn.

The Hamiltonian (1a) conserves the number of electron $N^e = \int \phi_a^* \phi_a dx$, thus we may go over to first quantization formalism and equivalently study the following Hamiltonian

$$h = -i \sum_{i=1}^{N^e} \frac{\partial}{\partial x^i} + J \sum_{i=1}^{N^e} \delta(x_i) \vec{\sigma}_i \quad (1.b)$$

The Hamiltonian now bears some similarity to a well-known model in quantum field theory called the chiral Gross-Neveu model [4].

To bring out the similarity [6] we introduce a new quantum number $\alpha = 1$ or 0 . α will be referred to as purity and the N^e electrons each carries $\alpha=1$ and the impurity $\alpha=0$, and we may rewrite h as follows ($N = N^e+1$)

$$h = -i \sum_{i=1}^{N=1} \alpha_i \partial_i + J \sum_{i,j=1}^N \delta(x_i - x_j) \vec{\sigma}_i \cdot \vec{\sigma}_j (\alpha_i - \alpha_j)^2 \quad (1.c)$$

The only difference with the former h is that now the impurity (with x^i that corresponds to $\alpha^i=0$) is allowed to move. But as it carries no kinetic energy we may form perfect wave packets that do not disperse and construct a localized impurity state. The factor $(\alpha_i - \alpha_j)^2$ insures that only electrons and impurity interact. We, then, want to solve the eigenvalue problem

$$h \mathcal{F}(x_i, \alpha_i, a_i) = E \mathcal{F}(x_i, \alpha_i, a_i)$$

where $\mathcal{F}(x, \alpha, a)$ is the wave function, depending on all the particles' coordinates. It is antisymmetrical since we are dealing with fermions:

$$\mathcal{F}(\dots x_i \alpha_i a_i \dots x_j \alpha_j a_j \dots) = - \mathcal{F}(\dots x_j \alpha_j a_j \dots x_i \alpha_i a_i \dots)$$

Since the Hamiltonian is invariant under spin rotations we are able to factorize the wave function $\mathcal{F} = F(x\alpha)\xi(a)$ and characterize the spin wave functions $\xi(a)$ by a Young Tableau R having up to two rows $R = [N-M, M]$ where $N = N^e + 1$ is the number of particles in the system and M the number of down spins. R completely characterizes the properties of the wave functions under the rotation group and under the permutation group S_N .

We turn now to the dynamics and show that h indeed can be completely diagonalized. We shall use the Bethe Ansatz [8] technique as brilliantly developed by Yang [9] and Gaudin [10]. Thus, divide configuration space into $N!$ regions labeled by permutation $Q \in S_N$. In the interior of region Q , defined by $0 < x_{Q_1} < \dots < x_{Q_N} < L$ the particles are free and we write F as a superposition of plane waves labeled by N momenta K_i and purity indices β_i , $[\beta_i=1, i=1\dots N^e, \beta_N=0]$

$$F(x, \alpha) = \sum_{p \in S_N} \xi_p(Q) \exp \left[i \sum_{j=1}^N K_{p_j} \cdot x_{Q_j} \right] \prod_{\alpha_{p_j}} \delta_{\alpha_{p_j}, \beta_{Q_j}} \quad (2)$$

The corresponding energy eigenvalue is obviously

$$E = \sum \beta_i K_i = \sum_{i=1}^{N^e} K_i^{(\text{electrons})} \quad (3)$$

The $N! \times N!$ matrix of constant coefficients $\xi_p(Q)$ is determined by matching the solutions across the boundaries. For example, consider the case of one electron and one impurity, $\beta_1=1, \beta_2=0$.

The wave function in the region I (i.e., $x_1 < x_2$) is given by

$$F_I = \xi_I(I) e^{i(K_1 x_1 + K_2 x_2)} \delta_{\alpha_1, 1} \delta_{\alpha_2, 0} + \xi_{p^{12}}(I) e^{i(K_2 x_1 + K_1 x_2)} \delta_{\alpha_2, 1} \delta_{\alpha_1, 0}$$

and in the region p^{12} i.e., $(x_2 < x_1)$ F is given by

$$F_{p^{12}} = \xi_I(p^{12}) e^{i(K_1 x_2 + K_2 x_1)} \delta_{\alpha_1, 0} \delta_{\alpha_2, 1} + \xi_{p^{12}}(p^{12}) e^{i(K_2 x_2 + K_1 x_1)} \delta_{\alpha_2, 0} \delta_{\alpha_1, 1}.$$

From the Hamiltonian h , we obtain the a 2×2 matrix Y relating the coefficients,

$$\begin{bmatrix} \xi_I(I) \\ \xi_{p^{12}}(I) \end{bmatrix} = Y \begin{bmatrix} \xi_I(p^{12}) \\ \xi_{p^{12}}(p^{12}) \end{bmatrix}$$

This can, obviously, always be done. The crucial test arises when we consider three particles with $3!=6$ regions not all of which are now adjacent. The regions can be arranged as shown in Figure 1.

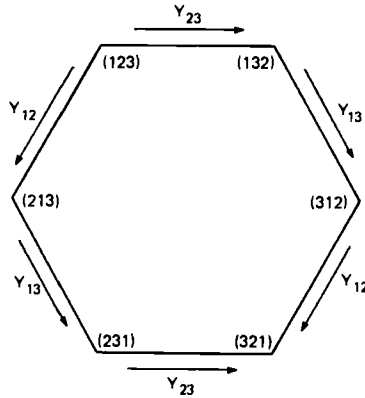


Figure 1: Schematic representation of the Y matrices connecting different regions in the case of 3 particles.

where the region $Q = \begin{pmatrix} 123 \\ 312 \end{pmatrix}$, for example, is denoted by (312) . Thus to go to region (312) from region (132) which is adjacent to it one needs to use the matrix Y_{13} which flips particles 1 and 3. When one deduces from the Hamiltonian h the set of matrices Y one may think that the problem is solved. This is rarely the case, as the matrices have to satisfy stringent consistency conditions. Thus, column $\xi_{(321)}$ can be obtained from column $\xi_{(123)}$ by repeated action of the Y -matrices

$$\xi_{(321)} = Y_{12} Y_{13} Y_{23} \xi_{(123)} \quad ,$$

but equally well it could have been obtained by a different path

$$\xi_{(321)} = Y_{23} Y_{13} Y_{12} \xi_{(123)} \quad .$$

Only if we get the same answer can we claim that the Ansatz (2) is consistent. For N particles one may think that one needs many more consistency conditions, but it turns out that the following are sufficient

$$Y_{ij} = Y_{ji} \quad , \quad Y_{ij}Y_{kl} = Y_{kl}Y_{ij} \quad (i \neq j \neq k \neq l \text{ are different})$$

$$Y_{ij}Y_{ik}Y_{jk} = Y_{jk}Y_{ik}Y_{ij}$$

The crucial point is that the Y matrices deduced from the Hamiltonian h (eq. (1b) and (1c)) do indeed satisfy the (factorization) condition (4) and we thus find that the Kondo Hamiltonian is soluble.

As all columns ξ_p can be consistently derived from any one of them we may concentrate on a reference column ξ_1 and determine it. To do so we need to impose boundary conditions which we choose to be periodic. Thus we require

$$F(\dots x_i = 0 \dots) = F(\dots x_i = L \dots) \quad (5)$$

Let me illustrate the procedure in an extremely simple case:

$$h = -i \frac{\partial}{\partial x} + 2J\delta(x)$$

and we try a solution

$$F(x) = [A\theta(-x) + B\theta(x)]e^{ikx}$$

Integrating across the boundary $x=0$, we find,

$$-i(B-A) + J(B+A) = 0 \quad .$$

so $\frac{B}{A} = \frac{i+J}{i-J} \equiv e^{i\phi}$ is the 1×1 matrix Y ,

relating the coefficients in the two regions. Note that no reflected (diffracted) wave is produced. This is an example of no-production phenomenon, characteristic of Bethe Ansatz systems: namely the same K values appear in both regions. If we had a second derivative this Ansatz would not be so consistent and a wave with $(-K)$ would be necessary.

To determine the allowed K 's we impose periodic boundary conditions, and find $e^{i\phi} = e^{ikL}$ so that the energy eigenvalue is given by $E = K = \frac{2\pi}{L} n + \frac{\phi}{L}$.

The effect of the interaction is thus given by the phase ϕ which shifts the energy from its free value.

We shall now write down the answer for the allowed K 's of the electrons of Hamiltonian (1). It is found, using Yang's method, that

$$K_i^{(\text{electron})} = \frac{2\pi}{L} n_i + \sum_{\gamma=1}^M [\theta(2\Lambda_\gamma - 2) - \pi] \quad (6)$$

where $\theta(x) = -2 \tan^{-1} \frac{x}{c}$, $c = \frac{2J}{1 - \frac{3}{4}J^2}$.

$\Lambda^1 \dots \Lambda^M$ are M complex numbers found by solving the following set of equations

$$N^e \theta(2\Lambda_\gamma - 2) + \theta(2\Lambda_\gamma) = \sum_{\delta=1}^M \theta(\Lambda_\gamma - \Lambda_\delta) + 2\pi I_\gamma. \quad (7)$$

The integers $\{n_j, I_\gamma\}$ are the quantum numbers of the eigenstates and each allowed choice uniquely determines a state. We are interested in the limit where N , the number of particles, and M , the number of down spins, go to infinity. In this case one can find the density of solutions of equation (7) and easily determine the properties of the eigenstates (6).

The ground state for example is a singlet and given by solving eq. (7) with a consecutive configuration, $I_{\gamma+1} = I_\gamma + 1$.

The simplest excitation is a triplet and is given by generating two "holes" in the ground state configuration, i.e., there are γ_1 and γ_2 such that $I_{\gamma_1+1} = I_{\gamma_1+2}$, etc. Denoting by Λ^1 and Λ^2 the Λ 's corresponding to the omitted integers we find [6] the excitation energy is given by

$$\Delta E(\text{triplet}) = \frac{N^e}{L} \left[\tan^{-1} e^{\frac{\pi}{c}(\Lambda^1 - 1)} + \tan^{-1} e^{\frac{\pi}{c}(\Lambda^2 - 1)} \right]. \quad (8)$$

We note that there is no mass gap, as Λ^i can be arbitrarily large and negative. This way one may go on and construct any number of elementary excitations.

We shall turn now to study the effect of a magnetic field on the system. We add to \mathcal{H} a magnetic term

$$\mathcal{H}_{\text{mag}} = -\mu H [S_z + \int \phi_a^*(x) (\sigma_z)_{ab} \phi_b(x)] \quad (9)$$

Since \mathcal{H} and \mathcal{H}_{mag} commute, the total Hamiltonian $\mathcal{H}_T \equiv \mathcal{H} + \mathcal{H}_{\text{mag}}$ has the same eigenstates as before. However, the ground state in the presence of a magnetic field will have finite magnetization. Indeed, each spin aligning with the magnetic field will gain magnetic energy given by $\Delta E = 2\mu H$ while it will lose an amount given by eq. (8). Thus the impurity magnetization M^i will be determined by these competing processes. It is found to be

$$M^i = \begin{cases} \frac{\mu}{\pi^{\frac{1}{2}}} \sum_{K=0}^{\infty} (-1)^K (K!)^{-1} (K+\frac{1}{2})^{(K-\frac{1}{2})} \exp[-(K+\frac{1}{2})] \left\{ \sqrt{\frac{e}{2\pi}} \frac{H}{T_0} \right\}^{2K+1}, & \sqrt{\frac{e}{2\pi}} \frac{H}{T_0} \leq 1 \\ \mu \left\{ 1 - \frac{\mu}{\mu^{3/2}} \int_0^{\infty} \frac{dt}{t} \sin(\pi t) \left[\sqrt{\frac{e}{2\pi}} \frac{H}{T_0} \right]^{-t} \left[\exp -t(\ln t - 1) \right] \Gamma(t+\frac{1}{2}) \right\}, & 1 \leq \sqrt{\frac{e}{2\pi}} \frac{H}{T_0}, (H \ll D) \end{cases} \quad (10)$$

We note that M^i has the advertised behavior, namely asymptotic freedom for high magnetic fields and a crossover to strong coupling regime when the field is decreased. For large fields, $H \gg T_H$, it approaches the free value

$$M^i \rightarrow \mu \left[1 - \frac{1}{2} \frac{1}{\log \left[\frac{H}{T_H} \right]} - \frac{1}{4} \frac{\log \log \left[\frac{H}{T_H} \right]}{\log^2 \left[\frac{H}{T_H} \right]} + \left[\frac{1}{\log \left[\frac{H}{T_H} \right]} \right]^3 \right] \quad (11)$$

and we have determined a new scale T_H by absorbing the $(\log)^{-2}$ term into the $(\log)^{-1}$ term. When $H \rightarrow 0$, on the other hand, we see that M^i approaches zero with a finite slope (i.e., susceptibility)

$$M^i \rightarrow \frac{\mu}{\pi T_0} H, \quad T_0 = \frac{N e}{L} e^{-\frac{\pi}{C}} \quad (12)$$

indicating screening of the impurity spin. The ratio between T_H and T_0 is a universal number and is given by

$$\frac{T_H}{T_0} = \sqrt{\frac{\pi}{e}}. \quad (13)$$

We have, then, achieved our goal to calculate a universal number and may have rested here had not Wilson chosen to calculate a different ratio T_K/T_0 ; but it is actually easy to convert to his number as both T_K and T_H characterize regions where perturbation theory is valid so that the ratio can be easily calculated. We find $\frac{T_K}{T_H} = 2\beta\gamma e^{-7/4}$ where $\ln \beta = \int_0^1 dx (1-x)^2 x (\pi^2 \csc^2 \pi x - x^{-2})$ and $\ln \gamma =$ Euler's constant.

Combining it with (13) we finally have

$$W = \frac{T_K}{T_0} = 2\beta\gamma\pi^{1/2} e^{-9/4} = 0.102676 \times 4\pi$$

which is in agreement with the value found numerically by Wilson

$$\frac{W}{4\pi} = 0.1032 \pm 0.0005 \quad .$$

To summarize:

The Kondo model can be completely diagonalized and its essential physics exposed. One can calculate various universal numbers analytically using our construction and find them to be in agreement with other ways of constructing the model [1]. Of course, using the exact solution one has access to new and even more interesting physics.

I would like to thank J. H. Lowenstein for a close and fruitful collaboration.

REFERENCES:

- [1] Wilson, K., Rev. Mod. Phys. 47, 773 (1975).
- [2] Anderson, P. W., Yuval, G., and Hamann, D. R., Phys. Rev. B1, 4464 (1970).
- [3] Gross, D., and Wilczek, F., Phys. Rev. Lett. 30, 1343 (1973); Politzer, H. D., Phys. Rev. Lett. 30, 1346 (1973).
- [4] Gross, D., and Neveu, A., Phys. Rev. D 10, 3235 (1979).
- [5] Andrei, N., and Lowenstein, J. H., Phys. Rev. Lett. 46, 356 (1981).
- [6] Andrei, N., Phys. Rev. Lett. 45, 379 (1980).
- [7] Wiegmann, P. B., Pis'ma Zh. Eksp. Theor. Fiz. 31, 392 (1980).
- [8] Bethe, H., Z. Phys. 71, 205 (1931).
- [9] Yang, C. N., Phys. Rev. Lett. 19, 1312 (1967).
- [10] Gaudin, M., Phys. Lett. 24A, 55 (1967).

This Page Intentionally Left Blank

KINKS OF FRACTIONAL CHARGE IN QUASI-ONE-DIMENSIONAL SYSTEMS

J. R. Schrieffer

Department of Physics and
Institute for Theoretical Physics
University of California
Santa Barbara, CA 93106 USA

Recent theoretical studies have shown that in quasi-one-dimensional conductors having a Peierls distortion of commensurability index n (ratio of the distortion period to lattice spacing), there exist excitations whose charge is $Q = \pm 2e/n$ or $\pm 2e/n \pm e$. These excitations are kinks in the order parameter ψ describing the lattice distortion. In trans polyacetylene, $n = 2$ and $Q = 0, \pm e$ with spin $S = \frac{1}{2}, 0$ respectively. For $n = 3$ (e.g., TTF-TCNQ at 19Kb) $Q = \pm \frac{2}{3}e, \pm \frac{1}{3}e, \pm \frac{4}{3}e$ with spin $S = 0, \frac{1}{2}, 0$ respectively. Electronic states localized at the kink have energies in the Peierls gaps. Properties of these stable fractionally charged objects are discussed.

INTRODUCTION

Fractionally charged excitations have often been considered in the past. Recently, however, it has been shown how a theory containing only fundamental particles of integer charge Q can lead to excitations having charge q which is a fraction of Q . In a continuum model of charge density waves in one-dimensional metals it was shown by Rice, Bishop, Krumhansl, and Trullinger¹ that commensurability effects can stabilize kinks which carry a charge $q = \Delta\theta/\pi$, where $\Delta\theta = 2\pi/n$, with n being the commensurability, e.g., $n = 2$ for a dimerized system, $n = 3$ for a trimerized system.

Independently, Jackiw and Rebbi² studied a relativistic field theory model in one dimension in which a spinless Dirac field is coupled to a self-interacting Bose field having two degenerate mean field ground states. They found that the model possessed soliton excitations which correspond to a change in the number of Dirac fermions in the vicinity of the kink being $\frac{1}{2}$. A c-number Fermion state at zero energy was found whose wavefunction was localized about the soliton. Depending on whether this state is occupied or not, the solitons would carry Fermion number $\pm \frac{1}{2}$, corresponding to charge $\pm e/2$.

Independently, J. Hubbard³ studied a one-dimensional tight binding chain with very large on-site repulsions and weaker nearest neighbor interactions. In this limit, the problem splits into a set of spinless Fermions and a set of Heisenberg spins. Hubbard showed that for one electron per two sites on average (quarter filled band), an injected electron would split into two kinks, each of charge $-e/2$. A related problem will be discussed by M. J. Rice in the following paper.

Motivated by experiments on trans (CH)_x, Su, Schrieffer, and Heeger (SSH)⁴ studied a one-dimensional model of electrons hopping along a chain, with the hopping integral linearly modulated by lattice displacements. Similar studies were carried out by Rice.⁶ Because of the Peierls distortion, the mean field ground state is degenerate. For the half-filled band (one electron per site on average) appropriate to undoped (CH)_x, the ground state is twofold degenerate. In this model, the symmetry breaking (loss of inversion symmetry in the phonon coordinates) is dynamically generated by the electron-phonon coupling as opposed to the imposed broken symmetry in the relativistic model of Jackiw and Rebbi.

For the (CH)_x model, it was found that there are low energy excited states corresponding^x to solitons which act as moving walls separating domains having different ground states, A and B. For each isolated soliton, an electronic state occurs in the center of the Peierls gap with a wavefunction centered about the soliton, as in the Jackiw and Rebbi model.^{3,6} An important difference, however, is that in (CH)_x electrons of both spin orientation enter symmetrically in the Fermi sea so that the charge $q = \pm e/2$ is "spin masked," being doubled to $q = \pm e$. These states are of spin zero, while for (CH)_x there is an added charge state of the soliton, $q = 0$ with spin $\frac{1}{2}$. Thus, instead of fractional charge being directly observable in (CH)_x, the effect appears as an interchange of the conventional charge-spin relations of electronic excitations in solids; instead of $q = \pm e$, $S = \frac{1}{2}$ for conventional electrons and holes, one has $q = \pm e$, $S = 0$ for charged solitons and $q = 0$, $S = \frac{1}{2}$ for neutral solitons. This breaking of the charge-spin relations of the underlying electron field is characteristic of the soliton excitations in a degenerate ground state system.

The one-third filled band (trimerized chain) was studied by Su and Schrieffer⁷ using the SSH hamiltonian. A consequence of the threefold degeneracy of the ground state is the existence of two types of kinks, K_1 and K_2 , each having three charge states: $2e/3$, $-e/3$, $-4e/3$ for K_1 and $-2e/3$, $e/3$, and $4e/3$ for K_2 . The spin of the $\pm e/3$ kinks is $S = \frac{1}{2}$ while the other kinks have $S = 0$. Thus, spin no longer masks fractional charge for trimerized chains. Below I discuss the origin of these results derived in collaboration with W. P. Su.

TRIMERIZED CHAIN

We consider the model Hamiltonian

$$H = - \sum_{ns} [t_0 - \alpha(u_{n+1} - u_n)] (c_{n+1,s}^\dagger c_{n,s} + \text{H.c.}) + \frac{K}{2} \sum_n (u_{n+1} - u_n)^2 + \frac{K}{2} \sum_n \dot{u}_n^2 \quad (1)$$

where u_n is the displacement of the n^{th} unit from its equilibrium position, c_n is the π electron creation operator on the n^{th} unit and M is the mass of one unit of the chain. K is an effective spring constant describing an harmonic approximation to the σ bond energy.

For the perfectly trimerized chain, we write in the (adiabatic) mean field approximation the most general function having the translation symmetry $\Delta n = 3$,

$$u_n = u \cos\left(\frac{2\pi}{3} n - \theta\right) \quad (2)$$

In Fig. 1 the energy per site is plotted as a function of θ for three values of the amplitude $u = 0, 0.04\text{\AA}$ and 0.07\AA where, as an example, we have chosen the (CH)_x parameters $t_0 = 2.5 \text{ eV}$, $\alpha = 4.81 \text{ eV/\AA}$, $K = 17.4 \text{ eV/\AA}^2$. The energy is minimized by $\theta - \pi/6 = 0, \pm 2\pi/3, \pm 4\pi/3, \text{ mod } \pm 2\pi, \dots$ and $u \sim 0.07\text{\AA}$. As one can see, the minimum energy path to go between these minima is by changing the phase angle θ with the amplitude essentially fixed. This leads to a small barrier height of order Δ^3/t^2 and a soliton width $\xi_3 \sim (2t_0/\Delta)^2 a$. For the dimerized system, only the amplitude variable exists, and the condensation energy passes through zero at the center of the kink, giving a barrier height of order Δ^2/t_0 per site, and a soliton width $\xi_2 \sim (2t_0/\Delta) a$. Hence, the width of the trimerized soliton is larger than that of the dimerized system by the factor $(2t_0/\Delta) \gg 1$.

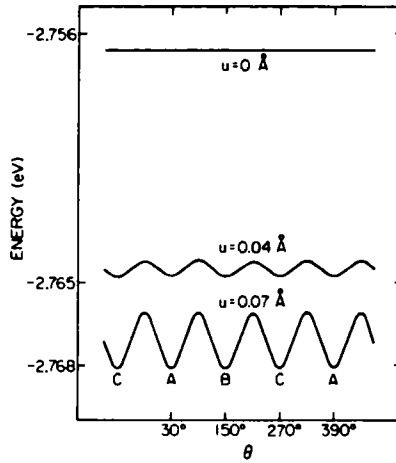


Figure 1: Total energy per site plotted as a function of the phase angle θ for three different values of the amplitude of trimerization u .

To determine the charge and spin of the kinks in the one-third filled band case, one cannot use the charge conjugation symmetry derivation which has been used for the dimerized system since this symmetry gives no significant information in this case. Rather, one can use 1) a Green function method to determine the charge and spin density of the soliton or, 2) a simple but general argument based on translational invariance and charge conservation. It is this latter argument we present below.

Consider an infinitely long chain, $-\infty < n < \infty$, with $\theta - \pi/6 = 0$, i.e., the A ground state. Suppose that θ_0 is adiabatically deformed so that the chain remains in the A state for $-\infty < n < n_1$, is in the B phase ($\theta - \pi/6 = 2\pi/3$) between n_1 and n_2 , is in the C phase ($\theta - \pi/6 = 2\pi$). The change of θ , near n_1 , n_2 , and n_3 need not be abrupt, but the transition region (soliton width) is assumed to be small compared to the spacings $n_2 - n_1$, and $n_3 - n_2$. There will be a certain width ξ_3 which will minimize the energy. As a result of this phase deformation, one has created 3 type I kinks (i.e., A \rightarrow B, B \rightarrow C, C \rightarrow A).

To determine the charge and spin of the kinks, suppose that one constructs two surfaces, one at n_0 far to the left of n_1 and another at n_r far to the right of n_3 . From global conservation of charge, we know that the total charge passing through these surfaces when θ_0 is slowly deformed is the negative of the sum of the charges of the three solitons, ΔQ ,

$$\int_{-\infty}^{\infty} [j_r(t) - j_l(t)] dt = -\Delta Q \quad (3)$$

where j_r and j_l are the electric current densities at the two surfaces. Since $\theta - \pi/6$ remains zero near n_0 , no current flows at this surface, while the charge passing the right-hand surface is the electronic charge in one unit cell of size $3a$ of the trimerized chain, i.e., $-2e$ and

$$\Delta Q = 2e \quad (4)$$

However, from the invariance of H under translations $\pm a$, it follows that all properties of the three kinks must be the same. Thus, the fundamental charge of type I kinks K_0 is

$$\Delta Q = 3Q_{K_0} = 2e \quad (5)$$

or

$$Q_{K_0} = \frac{2e}{3} . \quad (6)$$

A similar argument holds if the phase angle θ is decreased as one moves to the right so that one passes from A to C to B to A moving from left to right, with $\theta - \pi/6 = 0, -2\pi/3, -4\pi/3, -2\pi$. The electric current is now in the reverse direction, and the fundamental charge of these type II kinks \bar{K}_0 (antikinks) is

$$Q_{\bar{K}_0} = -\frac{2e}{3} . \quad (7)$$

Since spin up and spin down states are identically occupied during the adiabatic deformation, the spin of K_0 and \bar{K}_0 is zero,

$$S_{K_0} = S_{\bar{K}_0} = 0 . \quad (8)$$

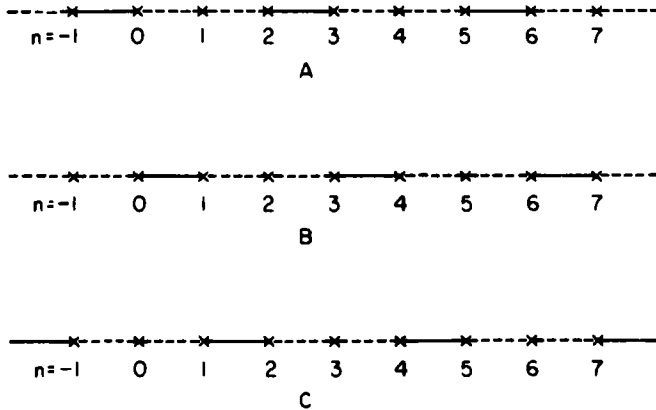


Figure 2: The three degenerate ground states of a perfectly trimerized chain.

As in the dimerized case, localized states are expected in the electronic energy gap connected with the solitons since states are missing from the valence band, leading to the fractional charge.

Using a Green function method one can determine the electronic spectrum in the presence of kinks. If one imposes a sharp kinks K or antikink \bar{K} on the system, one finds the spectrum shown in Fig. 3. For the kink, there are two gap states ϕ_ℓ and ϕ_u located in the upper half of the lower gap and symmetrically in the lower half of the upper gap. Direct calculation of the fractional parentage coefficients of these states shows that for K , ϕ_ℓ is derived $1/3$ from the bottom band and $2/3$ from the middle band, while ϕ_u comes $1/3$ from the top band and $2/3$ from the middle band. Since $1/3$ of a state is removed from the bottom band by K and all valence band states remain filled as K is created, it follows that if ϕ_ℓ and ϕ_u are unoccupied (K_0), then the charge of K_0 is $Q_{K_0} = 2/3 e$, in agreement with (6) and (8). However, if ϕ_ℓ is singly occupied then

$$Q_{K_1} = -\frac{e}{3}, \quad S_{K_1} = \frac{1}{2} \quad (9)$$

and if ϕ_ℓ is doubly occupied, then

$$Q_{K_2} = -\frac{4e}{3}, \quad S_{K_2} = 0. \quad (10)$$

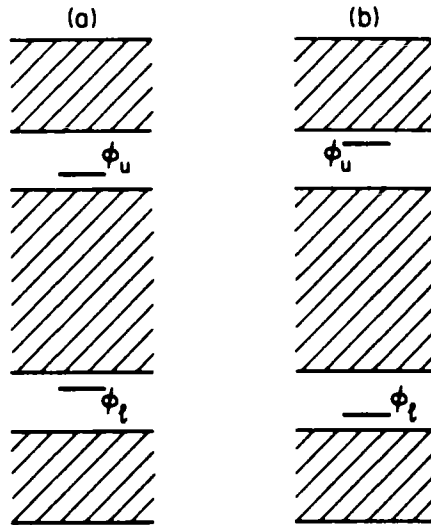


Figure 3: Gap states associated with (a) a sharp type-I kink and (b) a sharp type-II kink.

Similarly, the antikink \bar{K}_0 corresponds to having no holes in ϕ_0 , i.e., 2 electrons, so that its charge Q_s

$$Q_{\bar{K}_0} = -2e + 2\left(\frac{2e}{3}\right) = -\frac{2}{3}e, \quad (11)$$

in agreement with (7). In (11), the factor of two multiplying $2e/3$ arises from two spin orientations in accounting for the missing charge in the valence band due to the removal of $2/3$ of a state per spin by an antikink from the lowest band. Since all states are spin paired for K_0 , it follows that

$$S_{\bar{K}_0} = 0. \quad (12)$$

Finally, if ϕ_0 is singly occupied in \bar{K}_1 , then

$$Q_{\bar{K}_1} = +\frac{1}{3}e, \quad S_{\bar{K}_1} = \frac{1}{2} \quad (13)$$

and if ϕ_0 is empty, then

$$Q_{\bar{K}_2} = +\frac{4}{3}e, \quad S_{\bar{K}_2} = 0. \quad (14)$$

We see that K_v and \bar{K}_v act as antiparticles, having the same spin but reversed charge.

As in the dimerized case, there are constraints on creating solitons. If the chain is to be unaffected at infinity, topological constraints require that kinks must be created in pairs (KK or $\bar{K}\bar{K}$), triplets (KKK or $\bar{K}\bar{K}\bar{K}$), or combination of these allowed sets. From the kink quantum numbers (Q,S),

	K_v	\bar{K}_v
$v = 0$	(2/3,0)	(-2/3,0)
1	(-1/3, $\frac{1}{2}$)	(1/3, $\frac{1}{2}$)
2	(-4/3,0)	(4/3,0)

(15)

one can see that fractional charge and spin cannot be created globally but only locally. Hence, one would not be able to observe this type of fractional charge by measurements sampling the total charge of the system. Rather, the fractional charge reflects an internal distribution of an integer total charge of the system. One method for observing fractional charge in polymers is through the shot noise voltage fluctuation spectrum.

For n^{th} order commensurability, the primitive kink quantum numbers are

$$Q_{K_0} = \frac{2e}{n} = -Q_{\bar{K}_0} \quad \text{and} \quad S_{K_0} = S_{\bar{K}_0} = 0 \quad . \quad (16)$$

As above, K_0 and \bar{K}_0 can be decorated by electrons and holes in the localized states.

REAL OR APPARENT FRACTIONAL CHARGE?

The question has been raised whether the fractional charge discussed above is a quantum mechanically sharp observable or simply the quantum average of two or more integer charges, the result of which is fractional.⁸ In other words, if one can define a quantum operator Q_{op} which measures the charge of the soliton, then is the probability distribution $P(Q)$ of this operator in the one soliton sector a delta function at the fractional kink charge Q_K

$$P(Q) = \delta(Q - Q_K) \quad (17)$$

or is it given by

$$P(Q) = \alpha_1 \delta(Q - Q_1) + \alpha_2 \delta(Q - Q_2) \quad \dots \quad (18)$$

where Q_1, Q_2, \dots are integers, and only the expected charge is fractional

$$\langle Q_{op} \rangle = \sum_i \alpha_i Q_i = Q_K \quad ? \quad (19)$$

An example of apparent fractional charge, (18) and (19), is given by considering the H_2^+ ion. Suppose the electron is in the bonding (even parity) molecular orbital ψ_+ and the internuclear spacing is adiabatically increased to a very large value. In this limit ψ_+ reduces to a linear combination of 1s atomic orbitals ψ centered on the two protons ℓ and r ,

$$\psi_+ = \frac{1}{\sqrt{2}} \phi_\ell + \frac{1}{\sqrt{2}} \phi_r \quad . \quad (20)$$

If $Q_{op} = n_\ell$ is the occupation number for ϕ_ℓ , then

$$\langle \psi_+ | Q_{op} | \psi_+ \rangle = \frac{1}{2} \quad . \quad (21)$$

However, because Q_{op} has nonvanishing matrix elements between ϕ_+ and the nearly degenerate antibonding state ϕ_-

$$\psi_- = \frac{1}{\sqrt{2}} \phi_\ell - \frac{1}{\sqrt{2}} \phi_r \quad (22)$$

it follows that ψ_+ is not an eigenfunction of Q_{op} and the mean square fluctuation is given by

$$\langle \psi_+ | Q_{op}^2 | \psi_+ \rangle - \langle \psi_+ | Q_{op} | \psi_+ \rangle^2 = |\langle \psi_+ | Q_{op} | \psi_- \rangle|^2 = \frac{1}{4} . \quad (23)$$

For this simple example,

$$P(Q) = \frac{1}{2} \delta(Q) + \frac{1}{2} \delta(Q - 1) . \quad (24)$$

The usual quantum theory of measurement interpretation of this fictitious fractional charge is that while the average (expected) charge is fractional, only integer charge will be measured in any single experiment.

How is this related to the kinks? The essential point is that the off diagonal elements of the kink charge operator vanish exponentially in the limit of widely spaced kinks, rather than approaching a constant [$\frac{1}{2}$, see Eq. (23)] for the H_2 example. Therefore the one soliton state approaches an eigenfunction of Q_K for widely spaced kinks. This crucial difference arises from the fact that there is a unique electronic ground state of KK for any spacing, with no low lying excited states, all excited states having an energy of order or greater than Δ . Furthermore, even these states are not excited by a very long wavelength potential since a slowly varying potential only couples to the total charge of the system in the vicinity of the kink. Therefore, the fundamental fractional charge of the kink is sharp and arises totally from the depletion of states from the filled valence band when a soliton is formed.

If Q_{op} measures the up spin charge on the soliton, then for K_0 ,

$$P(Q) = \delta(Q - \frac{1}{2}) . \quad (24)$$

A totally separate but somewhat confusing question is the spacial distribution of charge located in the gap states. The discussion in the previous paragraph concerns configurations in which there are no electrons in the gap center states or when these states are completely filled. In this case there is no degeneracy, and a slowly varying field cannot admix excited states. Suppose, however, that we consider $K_1 \bar{K}_0$ so that there is one spin up electron in the gap center state of K . This configuration is degenerate with $K_0 \bar{K}_1$; however, for finite spacing there is mixing which produces the analog of the ψ_+ state of H

$$\frac{1}{\sqrt{2}} K_1 \bar{K}_0 + \frac{1}{\sqrt{2}} K_0 \bar{K}_1 . \quad (25)$$

In this case, the gap center electron fluctuates between K and \bar{K} leading to a fluctuating K charge whose probability distribution is

$$P_K(Q) = \frac{1}{2} \delta(Q - \frac{1}{2}) + \frac{1}{2} \delta(Q + \frac{1}{2}) , \quad (26)$$

which is to be compared with (24) showing a sharp distribution at $Q = \frac{1}{2}$. Clearly, the fractional charge associated with the vacuum deficit (or vacuum polarization) is a sharp quantum number, and it is this we refer to when discussing fractional charge. A fictitious fractional charge can enter from a totally different, and uninteresting effect, namely that analogous to charge fluctuations in H_2 .

CONCLUSIONS

While the noninteger charge effects discussed above will be screened by the dielectric constant of the medium, as in any solid the net charge macroscopically carried is the unscreened charge. The shot noise experiment should measure this charge. Fractionalized quantum numbers may be observable in other systems, such as domain walls in magnets and ^3He textures.⁹ Recently, Goldstone and Wilczek¹⁰ have shown how these ideas can be generalized to relativistic field theories in higher dimensions.

ACKNOWLEDGMENTS

The author is grateful to Drs. S. Kivelson and W. P. Su for helpful discussions. This work was supported in part by the National Science Foundation under Grant No. DMR80-07432, and Grant No. PHY77-27084.

REFERENCES

1. M. J. Rice, A. R. Bishop, J. A. Krumhansl, and S. E. Trullinger, *Phys. Rev. Lett.* **36**, 432 (1976).
2. R. Jackiw and C. Rebbi, *Phys. Rev. D.* **13**, 3398 (1976).
3. J. Hubbard, *Phys. Rev. B.* **17**, 494 (1978).
4. W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698 (1979); *Phys. Rev. B* **22**, 2099 (1980).
5. M. J. Rice, *Phys. Lett.* **A71**, 152 (1979).
6. For a review of the relation between fractionally charged kinds in condensed matter physics and relativistic field theory, see R. Jackiw and J. R. Schrieffer, *Nucl. Phys.* **B190** [FS 3], 253 (1981).
7. W. P. Su and J. R. Schrieffer, *Phys. Rev. Lett.* **46**, 738 (1981).
8. The author is indebted to Dr. A. K. Kerman and V. Weisskopf for raising this question. The discussion below relies on work of Dr. S. A. Kivelson and the author, to be published.
9. Jason Ho, private communication.
10. F. Wilczek and J. Goldstone, to be published.

This Page Intentionally Left Blank

THEORETICALLY PREDICTED DRUDE ABSORPTION BY A CONDUCTING CHARGED SOLITON IN DOPED-POLYACETYLENE

M. J. Rice

Xerox Webster Research Center
Webster, NY 14580, USA

With the aid of a novel quantum theory of the charged soliton we conclude that the observation of a "discrete Drude absorption" would constitute an experimental verification of charged-soliton transport in doped-polyacetylene.

The suggestion [1,2] that charged solitons may be generated in trans-polyacetylene $[(CH)_x]$ by light acceptor or donor doping has stimulated much subsequent experimental and theoretical interest in this subject.[3] A recent development of considerable interest has been the report by the Penn group [4] of a range of dopant concentration for which the homogeneously doped polymer exhibits metallic conductivity with an essentially zero Pauli spin susceptibility. Although electron transport arising from a small but finite Fermi-level density of localized one-electron states cannot be ruled out, [5] the conclusion that the metallic-like d.c. conductivity results from the transport of charged-solitons is seriously suggested for the first time. In order to conceive of an experiment that could distinguish between soliton and electron transport we have theoretically investigated the Drude absorption of a conducting charged-soliton in $(CH)_x$. Remarkably, we find that in addition to the usual Drude absorption expected for a diffusing free-carrier, the charged soliton's Drude absorption possesses a vibrational component which leads to a discrete absorption band at the frequency ω of the soliton's internal breathing mode.[6,7] This band is a consequence of the coupling of the internal and translational motions of the soliton which endows the mobile charged soliton with an oscillating electric dipole moment proportional to its mean velocity of translation. The observation of this "discrete Drude absorption" would constitute an experimental verification of charged-soliton transport in doped- $(CH)_x$.

In order to calculate the charged soliton's Drude absorption we employ the results of a quantum theory which we have developed for the charged soliton. The latter theory, which will be published in an independent article,[8] is a straightforward extension to quantum mechanics of a classical Hamiltonian theory of the soliton recently introduced by Rice and Mele.[6] In the quantum theory the soliton is described by the wavefunction $\psi(x, \ell)$ where $|\psi(x, \ell)|^2 dx d\ell$ determines the probability that the center of the soliton be found between the spatial points x and $x + dx$ with a length between ℓ and $\ell + d\ell$. The frequency-dependent conductivity of the free-charged soliton, $\sigma(\omega)$, is

$$\sigma(\omega) = \lim_{q \rightarrow 0} \left\{ -i\omega\Omega^{-1}e^2 \sum_{\alpha\gamma} P(E_\alpha) (2/\omega_{\gamma\alpha}) \times |v_{\gamma\alpha}(q)|^2 / (\omega_{\gamma\alpha}^2 - (\omega + i\delta)^2) \right\} \quad (1)$$

where $\{E_\alpha\}$ denote the energy eigenvalues of the free soliton, $\omega_{\gamma\alpha} = E_\gamma - E_\alpha$, Ω denotes the volume of the system, $P(E_\alpha) = \exp(-\beta E_\alpha) / \sum_\alpha \exp(-\beta E_\alpha)$ is the probability of finding the soliton in the state with energy E_α at temperature $T = 1/k_B\beta$, and $v_{\gamma\alpha}(q)$ is the matrix element

$$v_{\gamma\alpha}(q) = i \int_0^\infty d\ell \int_{-\infty}^\infty dx \psi_\gamma^*(x, \ell) v(q, \ell) \psi_\alpha(x, \ell) \quad (2)$$

of the velocity fluctuation operator

$$v(q, \ell) = [1/2M_S(\ell)] [p_x \exp(-iqx) + \exp(-iqx)p_x] \quad (3)$$

In (3) $M_S(\ell)$ denotes the classical length-dependent translational mass of the soliton [6] and $p_x = i\hbar \nabla_x$ is the soliton's translational momentum operator.

The energy eigenvalues of the free soliton are determined by the Schrödinger equation

$$H\Psi_\alpha(x, \ell) = E_\alpha \Psi_\alpha(x, \ell) \quad (4)$$

where the Hamiltonian operator H is

$$H = (1/4) (M_i(\ell)^{-1} p_\ell^2 + p_\ell^2 M_i(\ell)^{-1}) + V_i(\ell) + (1/2) M_S(\ell)^{-1} p_x^2 \quad (5)$$

Here, $M_i(\ell)$ denotes the classical length-dependent internal inertial mass of the soliton, [6] $p_\ell = i\hbar \nabla_\ell$ its internal momentum operator, and $V_i(\ell) = A\ell^{-1} + B\ell$ its internal potential energy. A and B are positive constants which determine the formation energy E_S and equilibrium length ℓ_0 of the classical static soliton according to the familiar relations $E_S = 2\sqrt{AB}$ and $\ell_0 = \sqrt{A/B}$. If we denote by M_S and M_i the translational and internal inertial masses of the latter soliton, we have according to the classical Hamiltonian theory, [6] $M_i(\ell) = M_i \ell_0 / \ell$ and $M_S(\ell) = M_S \ell_0 / \ell$. An extension of (4) and (5) to provide a quantal description of a charged soliton bound to an ionic dopant molecule will be discussed elsewhere.[8]

The eigenspectrum defined by Eqs. (4) and (5) and the latter stated ℓ -dependences of the classical masses, may be solved for exactly to yield the free-soliton eigenvalues

$$E_\alpha = E_{n,k} = [1 + (2/\kappa) (n + 1/2)] (M_S^2 c_0^4 + c_0^2 2k^2)^{1/2} \quad (6)$$

where $n = 0, 1, 2, \dots$, is an internal vibration quantum number and $k = (\pi/L)s$, with $s = \pm 1, \pm 2, \dots$, specifies the quantization of the soliton's translational momentum along a chain of length L ($L \rightarrow \infty$). $M_S c_0^2 = E_0$ and $\kappa/2$ denotes the dimensionless parameter $\kappa/2 = (2M_i \ell_0 A / \hbar^2)^{1/2}$. The latter's magnitude specifies the extent to which the nature of the internal motion of the soliton is quantum mechanical. Clearly, in the limit $\kappa \rightarrow \infty$, (6) yields the classical energy spectrum of the soliton. For (CH)_x we estimate $\kappa/2 \approx 3$. The corresponding eigenfunctions are

$$\psi_{n,k}(x, \ell) = L^{-1/2} \exp(-ikx) \psi_n(\ell/\lambda_k) \quad (7a)$$

where

$$\psi_n(y) = \exp(-y/2) y^{\kappa/2} \sum_{j=0}^n a_j y^j \quad (7b)$$

in which, for $j \neq 0$, the coefficients a_j are determined by the recurrence relation $a_{j+1} = -a_j(n-j)/[(j+1)(j+1+\kappa)]$ and a_0 by the normalization of $\psi_{n,k}(x, \ell)$. In (7a) λ_k denotes the characteristic length $\lambda_k = \ell_0 \kappa^{-1} [1 + (\kappa/M_S c_0)^2]^{-1/2}$. With the use of the relation $\kappa = 2E_S^0/\omega_i$, where $\omega_i = 2(B/2M_i \ell_0)^{1/2}$ is the classical harmonic frequency[6] of the internal breathing mode of the static soliton, (6) may be expanded for $2k^2/M_S \ll E_S^0$ to read

$$E_{n,k} = M_S c_0^2 + 2k^2/2M_{Sn} + (n+1/2) \omega_i \quad (8)$$

where $M_{Sn} = M_S \kappa/(\kappa+2n+1)$ may be regarded as the effective soliton translational mass in the vibrational sub-band n . We note that it follows from (7) that, for small k , the expectation value of the soliton length ℓ is just $\langle \ell \rangle = \ell_0 (\kappa+2n+1)/\kappa$ for the band n . This increase in $\langle \ell \rangle$ with vibrational excitation is responsible for the corresponding decrease in M_{Sn} . The remarkable contrast between the simple harmonic nature of the vibrational eigenspectrum of (8) and the decidedly anharmonic nature of the associated wavefunctions (7b) will be commented on in an independent publication.[8]

With the use of (8), (7) and the assumption that E_S^0 and $\omega_i \gg k_B T$, eq. (1) for $\sigma(\omega)$ may be evaluated to yield

$$\sigma(\omega) = \sigma_0 (1 - i\omega/\Gamma)^{-1} + (k_B T/\omega_i) \chi(\ell_{10}/\ell_{00})^2 (-i\omega/\Gamma) / (\omega_i^2 - \omega^2 - i\Gamma_i \omega) \quad (9)$$

where

$$\ell_{nn} = \int_0^\infty d\ell \psi_n(\ell) \ell \psi_n(\ell)$$

which, from eq. (7), gives $\ell_{10}/\ell_{00} = -(1+\kappa)^{-1/2}$. In evaluating (1) we have introduced a phenomenological constant Drude lifetime $\tau = \Gamma^{-1}$ and a natural width Γ_i for the internal breathing mode of the soliton. σ_0 is the d.c. conductivity $\sigma_0 = \Omega^{-1} e^2 \tau / M_{S0}$ and the first term of (9) corresponds to the Drude absorption ordinarily expected for a diffusing free-carrier. The second term is the discrete Drude absorption at ω_i arising from the coupling of the soliton's internal and translational motions. Its origin is clearly apparent from a classical interpretation of the soliton's velocity operator (3). Since, classically, ℓ is oscillating about its equilibrium value ℓ_0 , the velocity $v = p_x/M_S(\ell)$ of the soliton in a state of constant translational momentum p_x has an oscillatory component. As the soliton is charged it consequently is endowed with an oscillating electric dipole moment which is proportional to its mean velocity of translation. The integrated oscillator strength of the discrete Drude absorption, $S_i = S_0 (k_B T/\omega_i) (1+\kappa)^{-1}$, is therefore, understandably, proportional to the absolute temperature T . $S_0 = (\pi \Omega^{-1} e^2 / 2M_{S0})$ is the integrated oscillator strength of the ordinary Drude absorption.

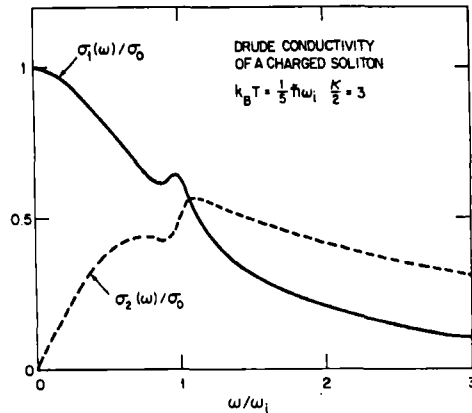


Figure 1: The real and imaginary parts of $\sigma(\omega)/\sigma_0$ vs ω/ω_i computed from eq. (9) for the parameter values stated in the text.

In Fig. 1 we have plotted the real and imaginary parts of $\sigma(\omega)/\sigma_0$ as a function of ω/ω_i for the representative choice of parameter values $\Gamma = \omega_i$, $\Gamma_i = \omega_i/5$, $k_B T = \omega_i/5$ and $\kappa = 2E_0/\omega_i = 6$. To date, an accurate microscopic calculation of ω_i for the charged soliton in $(CH)_x$ has not been undertaken. Following a phenomenological approach, Rice and Mele [6] have established the estimate 1280 cm^{-1} , whereas Su et al., [7] have arrived at the result $\omega_i = 780 \text{ cm}^{-1}$ on the basis of a simplified microscopic model of $(CH)_x$.

An experimental search for the discrete Drude absorption would be made difficult if the value of ω_i lies close to either of the two dopant-dependent intensely IR active modes already observed [9] for inhomogeneously doped $(CH)_x$ at 900 and 1370 cm^{-1} , respectively. The latter modes have been identified [10] as specific intrinsic structural vibrational excitations of the charged solitons.

It is hoped that this paper will stimulate an experimental search for this possible absorption band.

Finally, it is interesting to note that the relatively small value of $\kappa/2$ for $(CH)_x$ implies that the bond alternation amplitude of nature's simplest polymer has the character of a quantum field. The relevance of the quantum field theory literature [11] to the present problem is therefore called to attention.

We thank Gill Stengle, John Bardeen and Eugene Mele for several useful discussions.

REFERENCES

- [1] Rice, M. J., Phys. Lett. **71A**, 152 (1979).
- [2] Su, W. P., Schrieffer, J. R., and Heeger, A. J., Phys. Rev. Lett. **42**, 1698 (1979).
- [3] See, e.g., Proceedings of International Conference on Low-Dimensional Synthetic Metals (Helsingor, Denmark, August 10-15, 1980), to be published in *Chemica Scripta* (1981).

- [4] Ikekata, S., Kaufer, J., Woener, T., Pron, A., Druy, M. A., Sirak, A., Heeger, A. J., and MacDiarmid, A. G., Phys. Rev. Lett. 45, 1123 (1980): see, also, Peo, M., Roth, S. and Hocker, J., Proc. Inter. Conf. on Low-Dimensional Synthetic Metals (Helsingor, Denmark, Aug. 10-15, 1980) to be published in *Chemica Scripta* (1981); Epstein, A. J., Rommelmann, H., Druy, M. A., Heeger, A. J., and MacDiarmid, A. G., preprint 1980.
- [5] Mele, E. J., and Rice, M. J., Phys. Rev. B (submitted).
- [6] Rice, M. J., and Mele, E. J., Solid State Commun. 35, 487 (1980).
- [7] Su, W. P., Schrieffer, J. R., and Heeger, A. J., Phys. Rev. B 22, 2099 (1980).
- [8] Rice, M. J., and Mele, E. J., to be published.
- [9] Fincher, C. R., Jr., Ozaki, M., Heeger, A. J., and MacDiarmid, A. G., Phys. Rev. B 19, 4140 (1979).
- [10] Mele, E. J., and Rice, M. J., Phys. Rev. Lett 45, 926 (1980).
- [11] Jackiw, Rev. Mod. Phys. 49, 681 (1977).

This Page Intentionally Left Blank

POLARONS IN POLYACETYLENE

Alan R. Bishop
and
David K. Campbell

Theoretical Division and
Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545 USA

Recent theoretical studies of polyacetylene, $(\text{CH})_x$, have focussed on kink-like solitons in the trans isomer. It is shown that the same theoretical models also predict polaron-like solitons in both cis- and trans- $(\text{CH})_x$. The explicit form of these polarons, their relation to kinks, possible implications for experiment, and open questions are discussed.

INTRODUCTION

During the last few years the extensive theoretical and experimental interest in nonlinear excitations in polyacetylene $((\text{CH})_x)$ has focused almost exclusively on kink-like soliton states. [1-10] This is hardly surprising, for apart from their possible direct experimental implications [2,11] for transport properties, doping mechanisms, and the observed semi-conductor metal transition in $(\text{CH})_x$, these kink solitons, with their bizarre spin/charge assignments, [3] have stimulated theoretical work [12,13,14] on the existence and role of "fractional charge" in both solid state systems and field theory models.

Recently, however, it has been discovered [15,16,17] that the kink solitons are not the only nonlinear excitations predicted by theoretical models of $(\text{CH})_x$. Indeed, in both (numerical studies of) lattice models [15] and (analytic studies of) continuum theories, [16,17] a nonlinear "polaron" excitation has been found. Although more conventional in its properties than the kink soliton, the polaron, as we shall see, is the lowest energy excitation available to a single electron. Experimentally, the polaron may thus play an important role in the doping process or in electron injection and, theoretically, its presence further embellishes the already rich structure predicted for $(\text{CH})_x$. One particularly significant feature of the polaron is that it, unlike the kink soliton which can exist only in the trans isomer of $(\text{CH})_x$, is predicted to occur in both cis and trans isomers. The possible implications of this for comparative experimental studies of cis and trans are of great current interest. Thus in this article we shall focus on this "more conventional" but nonetheless interesting nonlinear excitation, describing its predicted nature in detail and discussing potential experimental implications briefly.

POLARONS IN TRANS- $(\text{CH})_x$

From the lattice model [3] for trans- $(\text{CH})_x$, a standard sequence of approximations [6] leads in the continuum limit to an effective adiabatic mean-field Hamiltonian in terms of the (real) gap parameter Δ and electron field Ψ . The result is

$$H = \int dy \left\{ \frac{\omega_Q^2}{g^2} \Delta^2(y) + \Psi^\dagger(y) \left[-iv_F \sigma_3 \frac{\partial}{\partial y} + \Delta(y) \sigma_1 \right] \Psi(y) \right\} \quad (1)$$

where ω_Q^2/g^2 is the net effective electron-phonon coupling constant, [3,6] σ_i is the i th Pauli matrix, $\Psi(y) = \begin{pmatrix} U(y) \\ V(y) \end{pmatrix}$, and v_F is the Fermi velocity. [18] For later purposes we note that in deriving this mean field Hamiltonian, the lattice kinetic energy - which would add a term proportional to $\Delta^2(y)$ in (1) - has been explicitly ignored. Obviously, this will have no effect on the static excitations we discuss below, but it will prove significant for on-going studies involving the dynamics of solitons in $(CH)_x$.

Variation of H leads to the single particle electron wave function equations

$$\varepsilon_n U_n(y) = -iv_F \frac{\partial}{\partial y} U_n(y) + \Delta(y) V_n(y) \quad (2)$$

$$\varepsilon_n V_n(y) = +iv_F \frac{\partial}{\partial y} V_n(y) + \Delta(y) U_n(y) \quad ,$$

and the self-consistent gap equation

$$\Delta(y) = -g^2(2\omega_Q^2)^{-1} \sum_{n,s} \left[V_n^*(y) U_n(y) + U_n^*(y) V_n(y) \right] \quad (3)$$

The summation in (3) is over occupied electron states and s is a spin label (suppressed in (1) and (2)).

Equations (2) and (3) are the continuum electron phonon equations for trans-(CH)_x, [5,6] and remarkably one can find analytic, closed form expressions for several nontrivial solutions. [5,6,16,17] Unsurprisingly, this is related to the (partial) soliton features of these equations. [16,17] Although we wish to focus on the polaron solutions to (2) and (3), for clarity and completeness we shall briefly review the other solutions.

First, the ground state of the infinite chain of trans-(CH)_x is two-fold degenerate (see Figure 1), and is described by $\Delta(y) = +\Delta_0$ or $\Delta(y) = -\Delta_0$ with [5,6]

$$\Delta_0 = W \exp(-\lambda^{-1}) \quad (4.a)$$

where

$$\lambda = \frac{g^2}{\pi v_F \omega_Q^2} \quad (4.b)$$

The parameters in (4) include W , the full one-electron bandwidth ($\cong 10$ eV is trans-(CH)_x); v_F , the Fermi velocity ($v_F = aW/2$, [18] where a is the underlying lattice spacing ($= 1.22$ Å in trans-(CH)_x); and, as indicated previously,

ω_0^2/g^2 , the net effective electron-phonon coupling constant (whose value is in effect determined by (4)). Since experimentally $\Delta_0 \approx 0.7$ eV, eqn. (4.a) implies the dimensionless coupling $\lambda \approx 0.4$.

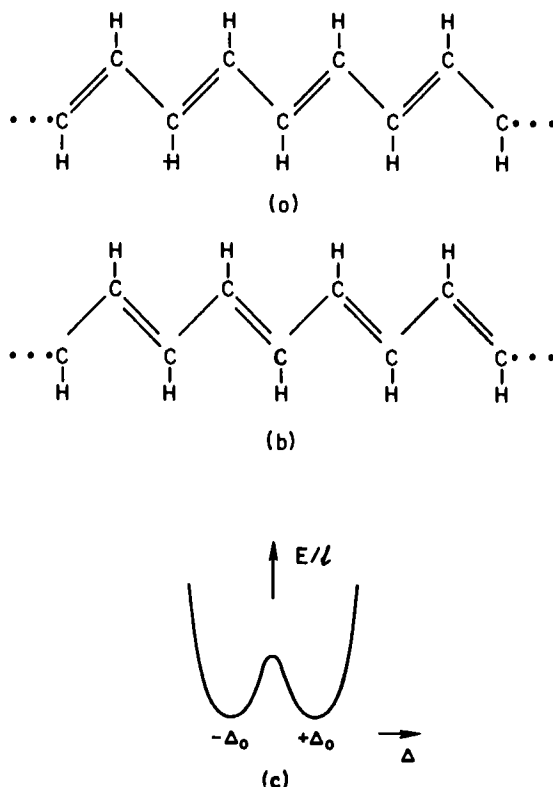


Figure 1: (a) and (b) - The two schematic bond structures corresponding to the two degenerate ground states for $\text{trans}-(\text{CH})_x$. (c) - A plot for $\text{trans}-(\text{CH})_x$ of the energy per unit length versus band gap parameter Δ for spatially constant Δ . Note the degeneracy between $\Delta = \pm \Delta_0$ and the local maximum (indicating instability) at $\Delta = 0$.

For constant Δ_0 , the single electron energy spectrum is given by $\varepsilon(k) = \pm \sqrt{k^2 v_F^2 + \Delta_0^2}$ (see Fig. 2) and the corresponding electron wave functions are simply plane waves, the explicit forms of which are given in refs. [5,6, and 17]. Physically, $\Delta_0 \neq 0$ implies a gap ($= 2\Delta_0$, see Fig. 2) around the fermi surface in the single electron spectrum; that this gap should exist in the continuum model of $(\text{CH})_x$ is a direct consequence of the well-known Peierls instability [19] of one-dimensional coupled electron-phonon systems. In chemical terms, $\Delta_0 \neq 0$ indicates a "bond alternation" between single and double bonds (see Fig. 1), and its value is proportional to the lengthening (shortening) of a single (double) bond with respect to the idealized uniform bond length which would occur were $\Delta_0 = 0$.

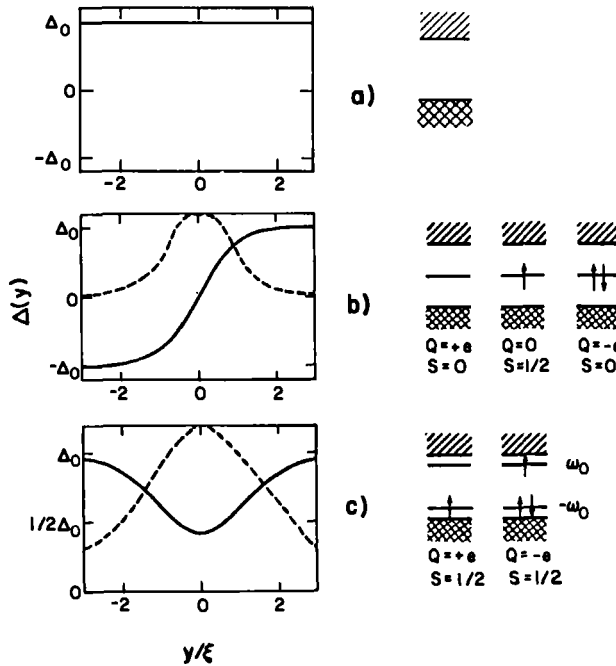


Figure 2: The spatial structure of the gap parameter ($\Delta(y)$) and the associated electronic levels for trans-(CH)_x for (a) the ground state; (b) the kink soliton; (c) the polaron. Solid lines in the left-hand drawings indicate $\Delta(y)$; dashed lines show electron densities for localized states. The right-hand drawings show the electronic levels in the conduction band (single shading), valence band (crossed shading), and gap states. Q = charge and S = spin.

The two-fold degeneracy of the ground state of trans-(CH)_x has the very important consequence that a localized, finite energy kink soliton - corresponding to a localized lattice "defect" - can exist; this kink (K) interpolates between the degenerate ground states (see Fig. 2) and has the explicit form

$$\Delta_K(y) = \Delta_0 \tanh v_F y / \Delta_0 \quad . \quad (5)$$

There is also an anti-kink (\bar{K}) soliton with $\Delta_{\bar{K}}(y) = -\Delta_K(y)$. The associated single electron energy levels consist of extended conduction ($\varepsilon(k) = +\sqrt{k^2 v_F^2 + \Delta_0^2}$) and valence ($\varepsilon(k) = -\sqrt{k^2 v_F^2 + \Delta_0^2}$) band states [5,6,17] which are modified (essentially "phase shifted") from the plane waves of the ground state and an additional localized "mid-gap" state (at $\varepsilon_0 = 0$, see Fig. 2). The explicit form of the wave function for this localized state is

$$U_0(y) = N_0 \operatorname{sech} \Delta_0 y / v_F \quad (6.a)$$

$$\text{and } V_0(y) = -iN_0 \operatorname{sech} \Delta_0 y / v_F \quad (6.b)$$

$$\text{with } N_0 = \sqrt{\frac{\Delta_0}{4v_F}}$$

For either K or \bar{K} , this "mid-gap" state can be occupied by 0, 1, or 2 electrons, leading to localized excitations with the bizarre spin/charge assignments [3,5,6,12] indicated in Fig. 2. It is these states that have excited the recent interest in "fractional charge." [12-14]

One vital consequence of the kink soliton's interpolating between the degenerate ground states is that there is a restriction - usually called a "topological constraint" [12,17] - on the production of kinks. Specifically, in the infinite chain polymer, kinks can be excited from the ground state only in pairs of kink and anti-kink. The interested reader should convince himself of this by showing that the production of a single kink from the ground state - that is, going from the $\Delta(y)$ configuration in Fig. 2a to that in Fig. 2b - requires overcoming an infinite (for an infinite length polymer) energy barrier.

Finally, we come to the polaron solution to equations (2) and (3). The recently discovered [16,17] analytic form for the gap parameter of the polaron can be written in the revealing form (c.f., eq. (5))

$$\Delta_P(y) = \Delta_0 - K_0 v_F \left\{ \tanh[K_0(y+y_0)] - \tanh[K_0(y-y_0)] \right\} \quad (7.a)$$

where

$$\tanh 2K_0 y_0 = K_0 v_F / \Delta_0 \quad (7.b)$$

This configuration for $\Delta(y)$, which corresponds to a spatially localized deviation from the ground state value (here chosen to be $+\Delta_0$), is sketched in Fig. 2c. Again, the single electron states include extended states in the conduction ($\epsilon(k) = +\sqrt{(kv_F)^2 + \Delta^2}$) and valence ($\epsilon(k) = -\sqrt{k^2 v_F^2 + \Delta^2}$) bands, which states are phase-shifted by the polaron "defect"; the explicit forms of the corresponding wave functions are given in ref. 17. In addition, two localized electronic states, with energies symmetrically placed, form in the gap at $\epsilon_{\pm} \equiv \pm \omega_0$ where

$$\omega_0 = \sqrt{\Delta_0^2 - (K_0 v_F)^2} \quad (8)$$

The electron wave functions for these localized states are, for $\epsilon = \pm \omega_0$

$$U_0(y) = N_0 \left\{ (1-i) \operatorname{sech} K_0(y+y_0) + (1+i) \operatorname{sech} K_0(y-y_0) \right\} \quad (9a)$$

and

$$V_0(y) = N_0 \{ (1+i) \operatorname{sech} K_0(y+y_0) + (1-i) \operatorname{sech} K_0(y-y_0) \} \quad (9b)$$

where $N_0 = \frac{1}{4}\sqrt{K_0}$, and for $\varepsilon = -\omega_0$ $U_{-0} \equiv U|_{\varepsilon=-\omega_0} = +iV_0$ and $V_{-0} \equiv V|_{\varepsilon=-\omega_0} = -iU_0$.

It is important to note that the polaron configuration for $\Delta(y)$ - eq. (6) - and the associated electron wave functions satisfy the electron part - eq. (2) - of the continuum equations for any $K_0 v_F$ in the allowed range $0 \leq K_0 v_F \leq \Delta_0$. It is the self-consistent gap condition - eq. (3) - which determines the specific value of $K_0 v_F$ for an actual solution to the coupled equations. Using some aspects of Soliton theory, [17] one can, in effect, convert the gap equation to a straightforward minimization problem. Apart from simplifying the problem technically, this is very appealing intuitively, for the quantity being minimized is essentially the energy of the full interacting electron-phonon system. In particular, one finds for the polaron-type configuration in trans-(CH)_x that

$$E_p^t(n_+, n_-, K_0) = (n_+ - n_- + 2)\omega_0 + \frac{4}{\pi} K_0 v_F - \frac{4}{\pi} \omega_0 \tan^{-1} \left(\frac{K_0 v_F}{\omega_0} \right) \quad (10)$$

where n_{\pm} are respectively the occupation numbers - which can be only 0, 1, or 2 - of the localized states $\varepsilon_+ = +\omega_0$ and $\varepsilon_- = -\omega_0$. Introducing θ such that $K_0 v_F = \Delta_0 \sin \theta$ and (see (8)) $\omega_0 = \Delta_0 \cos \theta$ and minimizing E_p^t with respect to θ give

$$0 = \sin \theta \left[\frac{4}{\pi} \theta - (n_+ - n_- + 2) \right], \quad 0 \leq \theta \leq \pi/2. \quad (11)$$

From eq. (11) it follows directly that for a stable, localized polaron solution the electrons must be distributed in one of two configurations:

(1) $n_+ = 1, n_- = 2$, the "electron polaron" state;

or

(2) $n_+ = 0, n_- = 1$, the "hole polaron" state.

Before describing these states in detail, we note that studying other possible configurations reveals some interesting phenomena. When $n_+ = n_-$ --- for $n_+ = 0, 1$, or 2 --- (11) implies $\theta = \pi/2$, so $K_0 v_F = \Delta_0$, $\omega_0 = 0$ (corresponding to the "mid-gap" state), and by (7.b) $y_0 \rightarrow \infty$. Hence these putative trans-(CH)_x "bi-polarons" separate into an infinitely separated kink/anti-kink pairs^x, with charges determined by the value of $n_{\pm} = n_-$. When $n_+ = 0, n_- = 2, \theta = 0$, so $\omega_0 = \Delta_0, K_0 v_F = 0$, and the "unoccupied" polaron collapses to the ground state, Δ_0 . Other choices of n_+ and n_- lead simply to combinations of the above excitations (infinitely separated in space); for example, $n_+ = 1, n_- = 0$ leads to a configuration of kink, anti-kink, and polaron.

Returning to the true, stable polaron states, we see that for $n_+ = 1, n_- = 2$ (electron polaron) or $n_+ = 0, n_- = 1$ (hole polaron) eq. (11) yields $\theta = \pi/4$ or

$$K_0 v_F = \Delta_0 / \sqrt{2} = \omega_0 \quad (12)$$

From eq. (10) the energy of this excitation is seen to be

$$E_p = \frac{2\sqrt{2}}{\pi} \Delta_0 \approx 0.90 \Delta_0 \quad (13)$$

From eqns. (6), (7.b), and (12) one sees that the spatial extent of the polaron ($\approx 11 \text{ \AA}$ for trans-(CH)_x) is slightly larger than that of the kink ($\approx 9 \text{ \AA}$ for trans-(CH)_x); this comparison is made visual in Fig. 2. Although general arguments suggesting the existence of polarons in systems like trans-(CH)_x - that is, (quasi)-one-dimensional Peierls dimerized chains - can be advanced, [20], it is encouraging to find that the continuum model, without any *ad hoc* additions, explicitly contains these excitations. It is also heartening that numerical simulations of the lattice model [15] have revealed a polaron excitation whose structure is in good agreement with that predicted by eq. (6), (7), and (12).

Another aspect of the polaron field configuration (eq. (7)) and corresponding energy expression (eq. (10)) is worth mentioning: namely, since the electron equations - eqns. (2) - are satisfied for all K_0 in the range $0 < K_0 < \Delta_0/v_F$, eq. (10) can be viewed as a single parameter (K_0) expression for the kink/anti-kink interaction energy as a function of separation, $2y_0$. Thus, for example, with $n_+ = n_-$ one finds, for $y_0 \gg \xi$,

$$(E_p^t(y_0) - E_p^t(\infty)) / E_p^t(\infty) \rightarrow \exp -2y_0/\xi \quad ,$$

with $\xi \equiv K_0^{-1}$, where K_0 changes with y according to eq. (7.5). This expression represents a slight improvement on previous continuum estimates of kink/anti-kink forces and is in good agreement with discrete simulations. [21]

Thus far we have treated the polaron as a static configuration, a solution to the adiabatic mean field Hamiltonian, H in eq. (1). For many experimental applications, it is vital to understand the dynamics of kink/anti-kink and kink-polaron interactions. A first step in this direction is to reinstate the lattice kinetic energy term, ΔH_{LKE} , explicitly dropped in deriving H . In the continuum limit this leads to an additional term of the form

$$\Delta H_{LKE} \equiv \frac{M}{2} \frac{1}{16\alpha^2} \int \frac{dx}{a} \left(\frac{\partial \Delta}{\partial t} \right)^2 \quad , \quad (14)$$

where M is the mass of a single CH unit ($\approx 13 m_p$, with m_p the proton's mass), a is the lattice spacing, and α is a constant whose value, in trans-(CH)_x is roughly 4.1 eV/Å. [3] Unfortunately, the (time-dependent) equations that follow from $H_D \equiv H + \Delta H_{LKE}$ have not been solved for non-trivially time-dependent Δ , and thus one can as yet say nothing definitive about kink and polaron dynamics in the continuum electron phonon models. However, some intriguing partial results in this area can be obtained in two ways. First, one can refer to the phenomenological ϕ^4 models [8,9] of trans-(CH)_x to discuss dynamics. Here one finds

that collisions between ϕ^4 kinks (and polarons [17,22]) have a fascinating and complex structure which suggests possibly interesting effects in transport and recombination processes in $(CH)_x$. Much of this structure in ϕ^4 appears due to a kink shape oscillation mode. [23] Since arguments for a similar mode in the continuum $(CH)_x$ theory have recently been advanced, [24] a similar complex dynamics might be expected. Second, for small velocity kinks and polarons, one can crudely estimate kinematic effects by simply making a Galilean boost of the static solution and retaining only lowest order, nonvanishing terms in the velocity; in effect, this gives an approximation to the soliton's inertial mass and thus gives some idea of its role in transport phenomena. For kink solitons, the results are given in refs. 3 and 6. For the polaron, replacing y by $\tilde{y} = y - v_p t$ in (6) and evaluating the expression in (14) gives

$$\Delta H_{LKE} = \frac{M}{2} \frac{1}{16\alpha^2} (K_0 v_F)^2 (K_0 v_p)^2 \quad (x)$$

$$\int_{-\infty}^{\infty} \frac{d\tilde{y}}{a} \left\{ \text{sech}^4 K_0 (\tilde{y} + y_0) + \text{sech}^4 K_0 (\tilde{y} - y_0) \right. \\ \left. - 2 \text{sech}^2 K_0 (\tilde{y} + y_0) \text{sech}^2 K_0 (\tilde{y} - y_0) \right\} \quad (15)$$

Evaluating the integrals (using (7.b) crucially) and defining M_p as the total coefficient of v_p^2 in (15) leads to

$$M_p = \frac{M}{16\alpha^2} \left(\frac{K_0}{a} \right) (K_0 v_F)^2 f(K_0 v_F) \quad (16a)$$

where

$$f(K_0 v_F) = \frac{8}{3} - \frac{4 w_0^2 \Delta_0}{(K_0 v_F)^3} \left\{ \ln \left(\frac{\Delta_0 + K_0 v_F}{\Delta_0 - K_0 v_F} \right) - \frac{2 K_0 v_F}{\Delta_0} \right\} \quad (16.b)$$

Using eq. (12), we find that $m_p \approx .9 m_e$, which shows that the polaron has an even smaller effective mass than the kink, for which $m_k \approx 5m_e$. [3,6]

POLARONS IN $\underline{cis}-(CH)_x$

From the perspective of the nonlinear excitations we are discussing, the most important difference between the trans and cis isomers of $(CH)_x$ is that for cis-(CH)_x, the two chemical structures corresponding to the putative ground state do not have the same energy. Thus there is a unique, nondegenerate ground state for cis-(CH)_x. This situation is shown in Figure 3. An immediate consequence is that since there are not two degenerate ground states for kink solitons to connect, there are no kink soliton solutions in cis-(CH)_x! A physically intuitive understanding of this situation follows by considering what would happen if one tried to create a kink/anti-kink pair from the ground state. Consider the case - relevant to cis-(CH)_x - where $\delta E/\ell$ is small (see Fig. 3) so the

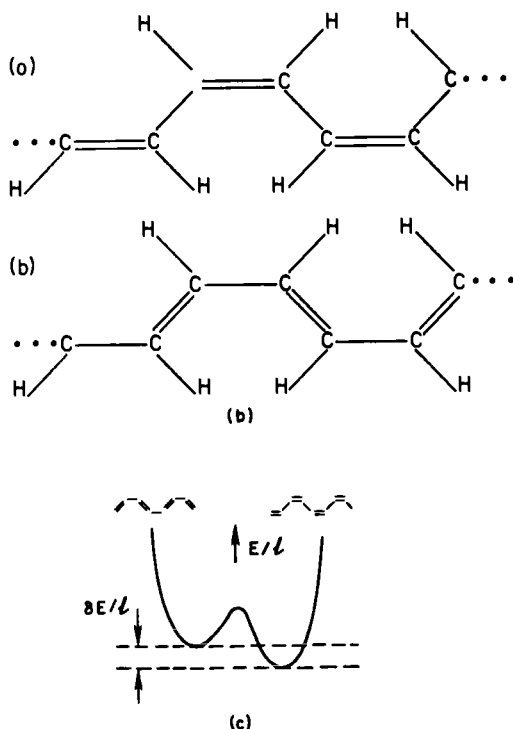


Figure 3: (a) and (b) - The two schematic band structures corresponding to the actual $\text{cis}-(\text{CH})_x$ ground state ((a), the cis-transoid configuration) and the metastable state ((b), the trans-cisoid configuration.) (c) - A plot for $\text{cis}-(\text{CH})_x$ of the energy per unit length versus band gap parameter Δ for spatially constant Δ . Note the unique ground state at $\Delta = \Delta_0$ and the metastable state of $\Delta = \Delta_{ms}$.

metastable and ground state have almost the same energy per unit length. One could then imagine constructing a configuration that interpolated between Δ_0 and Δ_{ms} (an anti-kink, say), remained in Δ_{ms} for a spatial distance d , and then interpolated between Δ_{ms} and Δ_0 (a kink). This configuration would thus look like eq. (6), with $d = \frac{m_s}{2y}$. For finite d , the configuration would have finite energy, but since $\delta E/l > 0$, for $d \rightarrow \infty$, its energy would become infinite like $(\delta E/l) \cdot d$. In this sense, one can say that the putative links in $\text{cis}-(\text{CH})_x$ are "confined." Of course, referring to eq. (6) we see that permanently confined KK pairs have the same structure as polarons. Thus, in $\text{cis}-(\text{CH})_x$, only polaron-type nonlinear excitations will exist.

To make this precise, it is very useful to study an explicit model of cis-(CH)_x (due to Brazovskii and Kirova [16]) which is both analytically solvable and related, in a certain limit, to the model we have examined for trans-(CH)_x. One assumes [16] that the gap parameter can be written as $\tilde{\Delta}(y) = \tilde{\Delta}_i(y) + \Delta_e$, where $\tilde{\Delta}_i(y)$ is sensitive to electron feedback (as for trans) but Δ_e is a constant, extrinsic component.[25] This ansatz can be motivated in terms of the effects arising from molecular orbitals other than the one included explicitly in eq. (1), (2), and (3). [16]

The mean field Hamiltonian for cis-(CH)_x then becomes [16]

$$H_{\text{cis}} = \int dy \left\{ \frac{w_0^2}{2} \Delta_i^2(y) + \Psi^\dagger(y) \left[-iv_F \sigma_3 \frac{\partial}{\partial y} + (\Delta_i(y) + \Delta_e) \sigma_1 \right] \Psi(y) \right\} \quad (17)$$

Although in the interest of simplicity we have not indicated this explicitly, it is very important to note that all the physical parameters in cis-(CH)_x - lattice spacing, effective electron phonon coupling, band width, fermi velocity, band gap - can have different values from those in trans. Comparing (1) and (17) we see that with the replacement $\Delta(y) \rightarrow \tilde{\Delta}(y) = \tilde{\Delta}_i(y) + \Delta_e$ the electron equations for trans are the same as for cis. Thus, mutatis mutandis, the structure of the electron spectrum and the eigenfunctions will be the same. The gap equation, however, is modified to read

$$\Delta_i(y) \equiv \tilde{\Delta}(y) - \Delta_e = -g^2(2w_0^2)^{-1} \sum_{n,s} [v_n^*(y)u_n(y) + u_n^*(y)v_n(y)] \quad (18)$$

which leads to different constraints on solutions and on bound state eigenvalues. Specifically, the ground state has the form $\tilde{\Delta}(y) = \tilde{\Delta}_0 = \Delta_i^0 + \Delta_e$ where

$$\Delta_i^0 \equiv w_c \exp(-\lambda_c^{-1}) \quad (19.a)$$

and

$$\Delta_e/\tilde{\Delta}_0 = \lambda_c \ln(\tilde{\Delta}_0/\Delta_i^0) \quad (19.b)$$

Here for clarity we have indicated explicitly that for cis-(CH)_x the dimensionless coupling, λ_c - see eq. (4.b) - and full band width, w_c , are not necessarily the same as for trans-(CH)_x. It can be shown directly that there are no kink solutions to the gap equation (18), in confirmation of our earlier intuitive arguments. For polaron configurations, however, we expect solutions, and indeed we find that exactly the same functional form - given by equations (7) - can satisfy the continuum equations for cis-(CH)_x, provided that K_0 (or w_0) is chosen appropriately. To determine K_0 we follow our previous procedure and again convert the gap equation to a minimization problem for the energy of the full interacting electron phonon system. For cis-(CH)_x the energy expression becomes

$$E_p^c(n_+, n_-, K_0) = (n_+ - n_- + 2)\omega_0 + \frac{4}{\pi} K_0 v_F - \frac{4}{\pi} \omega_0 \tan^{-1}(K_0 v_F / \omega_0) \\ + \frac{4}{\pi} \tilde{\Delta}_0 \gamma [\tanh^{-1}(K_0 v_F / \tilde{\Delta}_0) - K_0 v_F / \tilde{\Delta}_0] \quad , \quad (20)$$

where $\gamma = \Delta_0 / \lambda \tilde{\Delta}_0$. Again introducing θ such that $K_0 v_F = \tilde{\Delta}_0 \sin \theta$ and $\omega_0 = \tilde{\Delta}_0 \cos \theta$, we can write the equation minimizing E_p^c with respect to θ as

$$0 = \frac{4}{\pi} \sin \theta \left[\frac{\pi}{4} (n_+ - n_- + 2) - \theta - \gamma \tan \theta \right] \quad , \quad 0 \leq \theta \leq \pi/2 \quad . \quad (21)$$

In the limit of zero extrinsic gap, $\Delta_0 = 0$, $\gamma = 0$, and as expected eq. (21) reduces to the trans-(CH)_x result of eq. (11). For $\gamma > 0$, eq. (21) - unlike equation (11) - has solutions for all combinations of n_+ and n_- . These are summarized in Table I for the specific case of $\gamma = 1$.

Table I

n_+	n_-	N	Q	θ	$E_0 / \tilde{\Delta}_0$	INTERPRETATION
0	0	2	+2	0.71	1.43	BIPOLARON (h,h)
0	1	1	+1	0.38	0.98	POLARON (h)
0	2	0	0	0	0	GROUND STATE
1	0	3	+1	0.95	1.96	TRIPOLARON (h,h,e)
1	1	2	0	0.71	1.43	BIPOLARON (e-h)
1	2	1	-1	0.38	0.98	POLARON (e)
2	0	4	0	1.11	2.36	QUADRIPOLETON (e,e,h,h)
2	1	3	-1	0.95	1.96	TRIPOLARON (e,e,h)
2	2	2	-2	0.71	1.43	BIPOLARON (e,e)

The cis-(CH)_x polaron states for various values of $N = n_+ - n_- + 2$ for the case of $\gamma = 1$. $Q = 2 - n_- - n_+$ is the electric charge of the state, θ is the angle defined in the text (given here in radians), and $E_0 / \tilde{\Delta}_0$ gives the full energy of the excitation. The interpretation column gives the general nature of the excitation (polaron, bipolaron,...) and the type (electron \equiv e, hole \equiv h).

Of particular interest are the "bipolaron" solutions in which $n_+ = n_-$. In trans-(CH)_x the analogous solutions were infinitely separated KK pairs. Indeed, one can see from eq. (20) that γ essentially plays the role of a confinement parameter, leading to a term in the energy which increases linearly (recall eq. (7.b)) with the kink/anti-kink "separation," $2y_0$. This is the explicit realization of our earlier heuristic argument about kink confinement.

From Table I we see that both the polaron width ($2y_0$) and depth increase with increasing occupation number, $N \equiv n_+ + 2 - n_-$.

As in the case of trans, the problem of polaron dynamics in cis-(CH)_x is largely open. One can, of course, estimate the inertial mass as before, but the more interesting possible effects of collision dynamics - recombination, long-lived oscillatory states - have thus far eluded quantitative description.

IMPLICATIONS AND DISCUSSION

Since polarons have only recently been discovered [13,16,17] in the theoretical models for (CH)_x, the possible implications - in particular, for experiment - of their existence are not yet completely understood. Obviously the existence of polarons and kinks in trans-(CH)_x, contrasted with the existence of (multi-) polarons only in cis-(CH)_x, provides a natural distinction between these two isomers. To understand the quantitative implications of this distinction for experiment, however, further theoretical studies stressing the dynamics of kink and polaron interactions are needed. Two aspects of dynamics seem particularly crucial. First, within the time-dependent mean field approximation (see eqns. (1) and (14)), the nature of kink-kink and kink-polaron interactions must be understood. Second, one must go beyond the mean field approximation to study the full quantum dynamics of (CH)_x. Although there is, as yet, little quantitative progress in either of these areas, the recently noted connection between model relativistic field theories and the continuum theory of (CH)_x offer some qualitative suggestions. [17] First, this connection suggests that long-lived oscillatory states may exist in (CH)_x and that these may be important in (CH)_x dynamics. Second, known results on quantum dynamics in field theory raise the possibility that quantum fluctuations can destabilize the polaron in (CH)_x, causing it to disappear from the theory. Although preliminary quantum Monte Carlo studies [26] do show large quantum fluctuations in (CH)_x, analytic estimates [24] of the first order quantum corrections indicate that (to this order, at least) the polaron is not destabilized. Clearly, further quantitative work on these problem is required.

Despite the absence of critical quantitative detail, it is possible to indicate a number of interesting, potential experimental implications of polarons. First, on energy grounds [16,17] one knows that single electrons added - by doping or by injection - to cis- or trans-(CH)_x should form polarons. When many electrons are added, it is energetically favorable for them to form kink/anti-kink pairs (in trans) or multi-polarons (in cis). Thus for trans-(CH)_x, the very lightly doped material - an average of \leq one electron (or hole) per (CH)_x chain - might reflect polaron transport properties, whereas the more heavily doped material would exhibit (charged) kink transport behavior. In cis-(CH)_x, one would not expect this difference. Unfortunately this simple picture will be complicated by, among other features, the presence of neutral kinks "quenched" into undoped trans-(CH)_x samples and the uncertainties over isomerization from cis to trans upon doping. Nonetheless, if polarons are present in lightly doped cis- and trans-(CH)_x, then their optical absorption [27] is sufficiently different from that of kinks [28] that experiments comparing "mid-gap" and infrared absorption should be able to detect the difference. [27]

The distinction between allowed nonlinear excitations in cis and trans-(CH)_x suggests [16] a plausible scenario for the experimentally observed [29] contrasts between these isomers upon photoinjection (excitation of an electron-hole pair by intense radiation). In trans-(CH)_x, photoinjection leads to a photocurrent which has been interpreted [16,29] as resulting from a charged kink/anti-kink pair. In cis-(CH)_x, one observes instead a photoluminescence, suggesting the recombination of the e-h pair which could follow because of the effective confinement of the putative kinks (the localization of multi-polaron

states in *cis* aids recombination). Precisely the previously discussed theoretical dynamical studies are what is required to make this qualitative picture more precise.

In conclusion, we reiterate that recent theoretical developments [13,16,17] have shown that the nature of nonlinear excitations in polyacetylene may be even more intriguing and complex than previously believed. The central remaining challenge is to establish - or disprove - definitively the relevance of this elegant theoretical structure to the real material.

ACKNOWLEDGEMENTS

We are grateful to many colleagues for their advice and encouragement. Special thanks are due to J. L. Bredas, J. R. Schrieffer, and W. P. Su for enlightening discussions of their numerical simulations and to S. Etemad for his shared insights into the experimental situation.

REFERENCES

- [1] See for example, D. Bloor, Proc. Amer. Chem. Soc., Houston, April 1980; to appear in J. Chem. Phys.
- [2] For recent reviews see articles in Proc. Int. Conf. on Low-Dimensional Synthetic Metals (Helsingor, Denmark, August 1980) in *Chemica Scripta* 17 (1981); *Physics in One Dimension*, eds. J. Bernasconi and T. Schneider (Springer Verlag, 1981); and A. J. Heeger and A. G. MacDiarmid, p. 353-391 in *The Physics and Chemistry of Low Dimensional Solids*, ed. L. Alc  cer (Reidel, 1980).
- [3] Su, W. P., Schrieffer, J. R., and Heeger, A. J., Phys. Rev. Lett. 42, 1698 (1979); Phys. Rev. B 22, 2099 (1980).
- [4] Kotani, A., J. Phys. Soc. Japan 42, 408 and 416 (1977).
- [5] Brazovskii, S. A., JETP Letters 28, 606 (1978) (trans. of Pisma Zh. Eksp. Teor. Fiz. 28, 656 (1978)); and Soviet Phys. JETP 51, 342 (1980) (trans. of Zh. Eksp. Teor. Fiz. 78, 677 (1980)).
- [6] Takayama, H., Lin-Liu, Y. R. and Maki, K., Phys. Rev. B 21, 2388 (1980); Krumhansl, J. A., Horovitz, B., and Heeger, A. J., Solid State Commun. 34, 945 (1980); Horovitz, B., Solid State Commun. 34, 61 (1980).
- [7] Horovitz, B., Phys. Rev. Lett. 46, 742 (1981).
- [8] Rice, M. J., Phys. Lett. 71A, 152 (1979).
- [9] Rice, M. J., and Timonen, J., Phys. Lett. 73A, 368 (1979).
- [10] Mele, E. J., and Rice, M. J., in *Chemica Scripta* 17, 21 (1981).
- [11] See articles by Etemad S., and Rice, M. J., these proceedings.
- [12] Jackiw, R. and Rebbi, C., Phys. Rev. D 13, 3398 (1976); Jackiw, R., and Schrieffer, J. R., Nuc. Phys. B 190, 253 (1981).
- [13] Su, W. P., and Schrieffer, J. R., Phys. Rev. Lett. 46, 738 (1981).
- [14] Rice, M. J., and Mele, E. J., "Possibility of Solitons with charge $\pm e/2$ in Highly Correlated 1:2 Salts of TCNQ," Xerox Webster preprint, 1981.

- [15] Su, W. P., and Schrieffer, J. R., Proc. Nat. Acad. Sci. 77, 5526 (Physics) (1980); Bredas, J. L., Chance, R. R., and Silbey, R., Proceedings of International Conference on Low-Dimensional Conductors, Molecular Crystals and Liquid Crystals (Gordon Breach), to be published.
- [16] Brazovskii, S., and Kirova, N., Pisma Zh. Eksp. Teor. Fiz. 33, 6 (1981).
- [17] Campbell, D. K. and Bishop, A. R., Phys. Rev. B 24, 4859 (1981); Nuc. Phys. B (in press).
- [18] Following the standard conventions in the continuum model, we use units with $\hbar = 1$.
- [19] Peierls, R. E., Quantum Theory of Solids (Clarendon Press, Oxford, 1955) p. 108; Allender, D. Bray, J. W., and Bardeen, J., Phys. Rev. B. 9, 119 (1974).
- [20] Bishop, A. R., Solid State Commun. 33, 955 (1980).
- [21] Lin-Liu, Y. R., and Maki, K., Phys. Rev. B22 5754 (1980); Kivelson, S. unpublished.
- [22] The phenomenological ϕ^4 theories, if coupled to a phenomenological electron field describing only the localized, "gap" states, do contain both the kink solitons (of all charges) and the polaron.
- [23] Campbell, D. K., and Wingate, C., in preparation.
- [24] Nakahara, N., and Maki, K., preprint (1981).
- [25] In another context this ansatz has been discussed by Rice; see Rice, M. S., Phys. Rev. Lett. 37, 36 (1976).
- [26] Su, W. P., unpublished.
- [27] Fesser, K., Bishop, A. R., and Campbell, D. K., in preparation.
- [28] Gammel, J. T. and Krumhansl, J. A. Phys. Rev. B 24, 1035 (1981). Maki, K. and Nakahara, N., Phys. Rev. B 23, 5005 (1981); Horovitz, B., preprint (1981); Kivelson, S., Lee, T.-Y., Lin-Liu, Y. R., Peschel, I., and Yu, L., preprint (1981).
- [29] Etemad, S., these Proceedings.

LIGHT SCATTERING AND ABSORPTION IN POLYACETYLENE

Shahab Etemad and Alan J. Heeger

Department of Physics
University of Pennsylvania
Philadelphia, PA 19104, U.S.A.

Recent experimental results on the lattice dynamics, band structure and photoexcitations of polyacetylene are reviewed, and interpreted in terms of the soliton model. The data indicate that injection of an electron-electron (hole-hole) pair by chemical doping or photoinjection of an electron-hole pair invariably distort the lattice leading to formation of a soliton-antisoliton pair. It is concluded that unlike traditional semiconductors, the polyacetylene lattice is inherently unstable in the presence of electron and/or hole excitations.

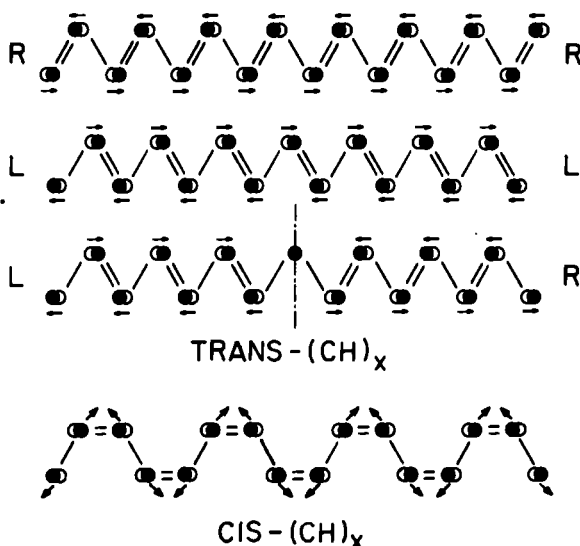
INTRODUCTION

Polyacetylene is the simplest conjugated polymer (1). It consists of weakly coupled chains of CH units forming a pseudo-one-dimensional lattice. Three of the four carbon valence electrons are in sp^2 hybridized orbitals; two of the σ -type bonds construct the 1-d lattice while the third forms a bond with the hydrogen side group. The 120° bond angle between these three electrons can be satisfied by two possible arrangements of the carbons; $\text{trans}-(\text{CH})_x$ and $\text{cis}-(\text{CH})_x$ with two and four CH monomer per unit cell respectively (see Fig. 1).

Figure 1

The lattice structure of $\text{trans}-(\text{CH})_x$ showing all the possible arrangements of the distortion pattern for a polymer chain.

The lattice structure of a $\text{cis}-(\text{CH})_x$ showing the distortion pattern for the nondegenerate cis-trans ground state.



In either isomer the remaining valence electron has the symmetry of a $2p_z$ orbital with its charge density lobes perpendicular to the plane defined by the other three. In terms of an energy-band description, the σ -bonds form low-lying completely filled bands, while the π -bond corresponds to a half-filled conduction band provided all the bond lengths are equivalent. As a result of Peierls instability of the 1-d metal (2), the bond lengths are not equivalent and alternate in size resulting in opening of an associated Peierls gap at the Fermi level.

Simple estimates lead to a picture of polyacetylene as a broad band pseudo-one-dimensional semiconductor. In order to describe the Peierls distorted state we consider, as an initial approximation, a model in which the π -electrons of $\text{trans}-(\text{CH})_x$ are treated in a tight binding approximation and the σ -electrons are assumed to move adiabatically with the nuclei. Let u_n be a configuration coordinate for displacement of the n th CH group along the molecular symmetry axis (x), where $u_n=0$ for the undimerized chain. The Hamiltonian is

$$H = - \sum_{hs} (t_{n+1,n} c_{n+1,s}^\dagger c_{n,s} + \text{h.c.}) + \sum_n \frac{K}{2} (u_{n+1} - u_n)^2 + \sum_n \frac{M}{2} \dot{u}_n^2 \quad [1]$$

where to first order in the u 's,

$$t_{n+1,n} = t_0 - \beta(u_{n+1} - u_n) \quad [2]$$

M is the mass of the CH unit, K is the spring constant for the σ -energy when expanded to second order about the equilibrium undimerized systems and c_{ns}^\dagger (c_{ns}) creates (annihilates) a π -electron of spin s on the n th CH group. The band structure of the perfect infinite dimerized trans structure is shown in

Figure 2
The band structure
of $\text{trans}-(\text{CH})_x$.

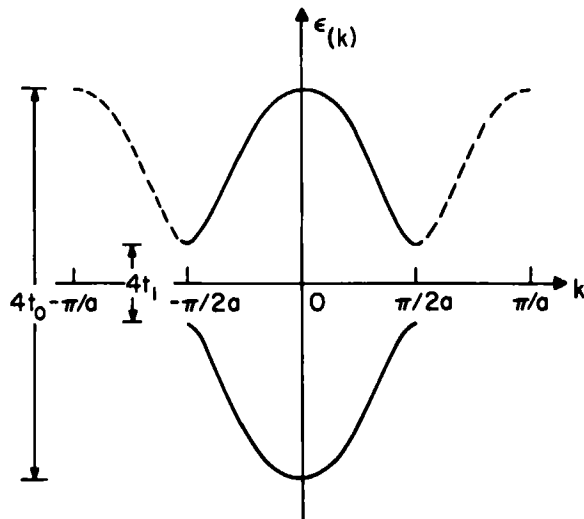


Fig. 2. In this perfect structure, the displacements are of the form

$$u_n = \pm (-1)^n u_0 \quad [3]$$

where \pm corresponds to the two possible degenerate structures (double bonds pointing "left" and double bonds pointing "right"). The transfer integrals for the perfect chain are

$$t_{n+1,n} = \begin{cases} t_0 - t_1 & \text{"single" bond} \\ t_0 + t_1 & \text{"double" bond} \end{cases} \quad [4]$$

The overall bandwidth is $4t_0 \sim 8-10$ eV; whereas the energy gap $2\Delta = 4t_1$, depends on the magnitude of the distortion. For example, in $\text{trans}-(\text{CH})_x$ for $2\Delta = 4t_1 \sim 1.5$ eV and $\beta = 6.9$ eV/Å, as inferred from the analysis of the optical absorption coefficient and the lattice dynamic results (see below), the value of u_0 which minimizes the ground state energy can be obtained from

$$\Delta = 4\beta u_0 \quad [5]$$

to be $u_0 \sim .025$ Å compared with an intersite separation of $a = 1.2$ Å along the chain direction. As a result of the large bandwidth and unsaturated π -system, $(\text{CH})_x$ is analogous to the traditional inorganic semiconductors. The transverse transfer integral t_\perp is, however, much less than t_0 . The large nearest neighbor interchain separation of ~ 4 Å (4) implies $t_\perp \sim .1$ eV compared to $t_0 \sim 2.5$ eV; thus justifying the pseudo-one-dimensionality of the system.

There has been a growing interest in the study of the physical properties of $(\text{CH})_x$ due to the fact that the structure of $\text{trans}-(\text{CH})_x$ exhibits a broken symmetry and has a two-fold degenerate ground state. (see Fig. 1) As a result, one expects soliton-like excitations, in the form of bond alternation domain walls, connecting the two degenerate phases. The properties of solitons in $(\text{CH})_x$ have been explored theoretically in several recent papers (5-8), which show that the coupling of these conformational excitations to the π -electrons leads to unusual electrical and magnetic properties. The possibility of experimental studies of such solitons in polyacetylene, therefore, represents a unique opportunity to explore nonlinear phenomena in condensed matter physics.

In an effort to characterize the soliton excitations of polyacetylene, Su, Scheieffer, and Heeger (SSH) considered the two domains with L domain to the left and R domain to the right of a domain wall or soliton as shown in Fig (1C). They determined the properties of the soliton (creation energy, width, mass, spin, etc.) in terms of the microscopic parameters of equation (1) by minimizing the ground-state energy of the system for a displacement pattern which reduces to the L and R phases as one moves to the far left or right. Using a staggered order parameter trial function ψ_n ,

$$\psi_n = (-1)^n u_0 \tanh(na/\xi) \quad [6]$$

they find that due to the competition between the elastic term (increases with a decrease in ξ) and the condensation energy (increases with an increase in ξ) the soliton in $\text{trans}-(\text{CH})_x$ is spread over approximately 15 lattice sites, i.e., $2\xi = 15$. Using the adiabatic approximation for the electronic motion, they estimate the translational mass of soliton; i.e., the inertia of the bond alternation pattern of equation 6 to motion along the chain, to be

$$m_s \approx \frac{4}{3} M \left(\frac{u_p^2}{a\xi} \right) \quad [7]$$

The enormous reduction in the soliton mass from the mass of a CH unit is due to the small amplitude of the bond alternation. Using the experimentally determined values of β and Δ (see below), we estimate $m_s \sim m_e$ where m_e is the electron mass. Since there is complete degeneracy (the soliton can be anywhere), and since the mass is small the soliton is expected to be highly mobile, but by topological constraints confined to move on a given chain (9,10).

Associated with the soliton is a localized electronic state situated at mid-gap; i.e., half-way between the bonding valence band and the antibonding conduction band. SSH have pointed out that the solitons in trans-(CH)_x are not the phase solitons which correspond to a 180° phase shift of the electronic-charge-density wave, rather they are amplitude solitons with reversed charge-spin relations (5). As a result, if the localized state contains one electron the soliton is neutral, with spin ½, and therefore is paramagnetic. On the other hand, if the localized state is empty (doubly occupied) the soliton is positively (negatively) charged, with spin 0, and non-magnetic. The soliton creation energy in trans-(CH)_x; i.e., energy to generate a mid-gap state, is estimated by SSH to be $E_s \sim .43$ eV which is less than half the gap, $\Delta \sim .7$ eV. The analytical equation for E_s has been found to be

$$2E_s = \frac{4}{\pi} \Delta < 2\Delta \quad [8]$$

using the continuum limit approximation to the SSH model (7). This result is of fundamental importance to the nature of the photoexcitations and the process of chemical doping in trans-(CH)_x. Equation (8) clearly states that (within the model described by equation 1, from energetic considerations trans-(CH)_x is unstable to the presence of an electron-hole (e-h) pair, or an e-e (h-h) pair in or near the conduction (valence) band, and distorts to form two mid-gap states of appropriate charge.

In this paper we review the results of the recent light scattering and absorption experiments in cis- and trans-(CH)_x. We show that the strong coupling of these nonlinear excitations to the electronic structure of the polymer leads directly to a number of unusual results which cannot be explained in terms of "linearized" concepts. The organization of this paper is as follows: We first summarize the results of a recent study of the lattice dynamics of trans-(CH)_x, including both the unperturbed polymer and the doped polymer containing charged solitons at dilute concentrations. The generation of solitons upon doping rearranges the bondlengths giving rise to antisymmetric displacement pattern about the soliton center. The lattice dynamics is locally perturbed, inducing infrared active modes which correspond to the oscillation of charge across soliton center. The detailed quantitative agreement between calculated and experimentally observed infrared active vibrational modes (IAVM) provides specific experimental evidence for doping through soliton generation in polyacetylene. The appearance of the IAVM is concurrent with generation of mid-gap state readily detectable in the absorption spectra of lightly doped polyacetylene. The experimental results can naturally be understood in terms of the soliton model; the presence of a soliton breaks the translation symmetry of the lattice, thereby removing oscillator strength from the interband optical transition and creating a mid-gap absorption band due to the optical transition between a band state and the soliton level. The observation of doping induced IAVM concurrent with the appearance of mid-gap states confirms the theoretical prediction that trans-(CH)_x lattice is unstable to the presence of e-e (h-h) pair in the conduction (valence) band.

To avoid possible complexities associated with soliton generation by chemical doping we have also considered experimental configurations that have permitted us to study the response of the lattice to direct photo-injection of e-h pairs. The existence of the two isomers of polyacetylene has provided the conceptual and experimental basis for the study of the photoexcitations in the undoped polymer in terms of photogeneration of solitons. The results of two complementary experiments; i.e., photoluminescence and photoconductivity in cis-(CH)_x and trans-(CH)_x are presented. For cis-(CH)_x, we find band edge luminescence in the scattered light spectrum near the interband absorption edge, but we are unable to detect any photoconductive response. Isomerization of the same sample quenches the luminescence signal, and gives rise to the appearance of a large photoconductive response. The spectral dependence of the photoconductivity in trans-(CH)_x can be interpreted in terms of charged solitons, photogenerated either directly (threshold $\hbar\omega = 4\Delta/\pi$) or indirectly through coupling of the lattice to electron-hole pair excitations ($\hbar\omega > 2\Delta$). The observation of luminescence in cis-(CH)_x, but not in trans-(CH)_x; and the observation of photoconductivity in trans-(CH)_x, but not in cis-(CH)_x provide confirmation of the proposal that solitons are the photogenerated carriers. In trans-(CH)_x, the degenerate ground state leads to free soliton excitations, absence of band edge luminescence, and photoconductivity. In cis-(CH)_x the non-degenerate ground state leads to confinement of the photogenerated carriers, absence of photoconductivity, and to the band edge luminescence.

IAVM OF CHARGED SOLITONS (12)

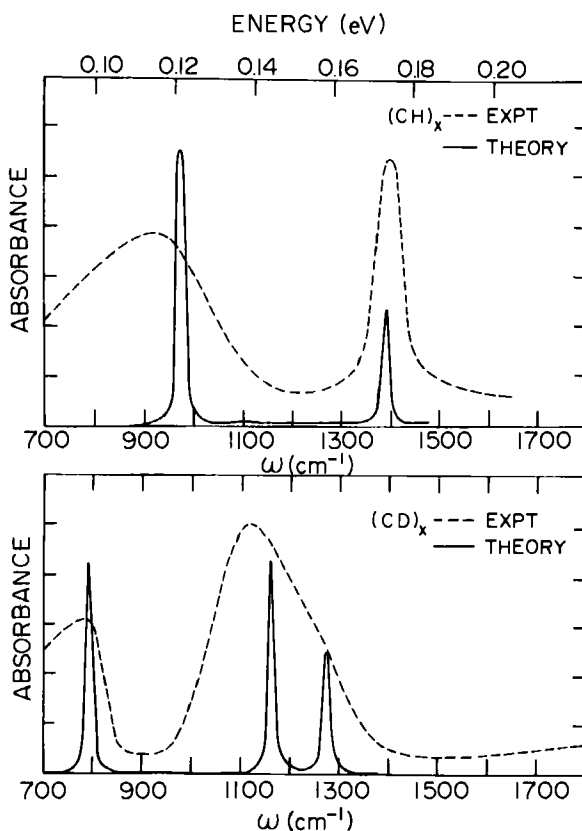
Specific evidence for the formation of charged solitons upon doping has emerged from infrared studies of the donor and acceptor states in lightly doped polyacetylene. In a series of experiments with various dopants in both (CH)_x (15) and (CD)_x (12) it was found that upon dilute doping new absorption modes appeared in the IR region, with remarkable intensity. The doping induced modes are primarily polarized parallel to the chain direction and their intensities grow in proportion to the dopant level, becoming comparable to any of the intrinsic IR lines of the undoped polymer at about 0.1%. Thus the doping induced IR modes have oscillator strengths enhanced by approximately 10^3 ; such a large enhancement must arise from coupling of the new vibrational modes (induced by doping) to the electronic oscillator strength of the polyene chain. These results are quite general; the same modes can be observed for iodine and AsF₅ (p-type), and for Na doping (n-type) (15). Thus these intense absorptions are not due to specific vibrations of the dopant molecules, nor to detailed interactions between the dopants and the polymer chain. The observed generality suggests that the intense IR absorption modes are intrinsic features of the doped (CH)_x and (CD)_x chains.

Mele and Rice have been able to successfully interpret these results in terms of a theory of the lattice dynamics of solitons in trans-(CH)_x or trans-(CD)_x (16-17). In the frequency range of the vibrational modes they have found several internal modes peculiar to the soliton structure. In particular, three of these modes were found to be strongly infrared active, deriving their oscillator strength from interactions with the π -electrons. Mele and Rice have shown that the dominant motions associated with the infrared active vibrational modes (IAVM) of a soliton involve an antisymmetric contraction of the single (or double) bonds on one side of the soliton center and an expansion on the other, thus driving charge back and forth across the soliton center. Due to the spatial spread of the soliton over about 15 sites, they find that the asymmetric oscillation of the electronic charge generates a large oscillating dipole moment consistent with the experimentally observed enhanced oscillator strength of the dopant induced IAVM.

The force field model constructed by Mele and Rice relied on a single adjustable parameter β which included the effect of nonlocal coupling of the atomic displacements through their interaction with the extended π -electronic states (16). The value of β was determined by a fit of the calculated zone center frequencies of gerade modes to the observed Raman lines in unperturbed trans-(CH)_x and (CD)_x (12). Despite considerable change in the Raman spectra of polyacetylene upon deuteration, the fit to the Raman lines in both (CH)_x and (CD)_x is obtained with the same value of $\beta=6.9 \text{ eV/\AA}$. These major changes which occur in the phonon spectrum on going from (CH)_x to (CD)_x arise since the increase in mass is sufficiently large compared to the mass of the carbon atom that non-trivial changes occur (i.e., beyond a simple scaling by $M^{-1/2}$) in the normal modes.

A comparison of the IAVM of solitons in trans-(CH)_x and trans-(CD)_x is shown

Figure 3
The comparison of the experimentally derived and the theoretically calculated additional absorption due to the infrared active vibrational modes (IAVM) of solitons in trans-(CH)_x and trans-(CD)_x .



in Fig 3. The dashed curves are the experimental results for a concentration of approximately 0.1% and the solid curves are the result of the calculations

with no adjustable parameter (the value of $\beta = 6.9 \text{ eV/\AA}$ was already fixed). Besides the large frequency shifts upon deuteration, the detailed structure of the soliton IAVM that is driven by the higher frequency stretching oscillations is also changed. First, the absorption is considerably broader and slightly asymmetric, with a shoulder on its high energy side. The width (measured as full width at half the peak amplitude) of this mode is changed by a factor of five, from about 50 cm^{-1} in $(\text{CH})_x$ to about 250 cm^{-1} in $(\text{CD})_x$. Secondly, the relative integrated oscillator strengths (Ω) of the two absorption modes is changed. The experimentally determined ratio in $(\text{CH})_x$, $\Omega(900)/\Omega(1370)$, is about 2/3, whereas the analogous ratio in $(\text{CD})_x$, $\Omega(780)/\Omega(1120)$ is about 3/2.

As shown in Fig. 3, the calculated infrared absorption associated with the IAVM of solitons is in remarkably good agreement with the experimental results. First, the frequencies are in agreement with the experimental results to an accuracy of a few percent for all the IAVM for both $(\text{CH})_x$ and $(\text{CD})_x$. Secondly, the increased width and asymmetry of the higher frequency mode in $(\text{CD})_x$ appears to be accounted for by the blue shifting of the two degenerate modes in the low frequency line of trans-($\text{CH})_x$. Finally as in the case for $(\text{CH})_x$, the calculated absolute integrated oscillator strengths correspond to $\sim 20\%$ of the total associated with the added charge, and they agree with the experimental values to within a factor of 2-3. The relative oscillator strengths of the different modes are even more accurate. In $(\text{CH})_x$, the lower frequency mode is more intense by about 3:2, whereas in $(\text{CD})_x$ we find that the higher frequency absorption is more intense by $\sim 3:2$. In both cases the experimental and theoretical intensity ratios are in close agreement.

The unusually precise account of the lattice dynamics of the unperturbed polymer and characteristic vibrations of the local deformation associated with the presence of solitons in both $(\text{CH})_x$ and $(\text{CD})_x$ is achieved by fixing the single adjustable parameter $\beta = 6.9 \text{ eV/\AA}$. As indicated above, this value of β implies a horizontal displacement of $\sim 0.025\text{\AA}$ for the CH unit, or $\sim 0.044\text{\AA}$ for the bond length change, with an energy gap of $2\Delta = 1.5\text{eV}$. Using eq. 7 the inferred value of the soliton mass is $m_s \sim m_e$; i.e., quantum effects associated with the soliton states may dominate its physical properties.

SOLITON MID-GAP STATE (11)

In this section we examine, from a band structure point of view, the response of trans-($\text{CH})_x$ lattice to the presence of an electron-electron (hole-hole) pair in or near the conduction(valence) band. In contrast to the case of conventional semiconductors, we find that donor (acceptor) doping does not result in generation of localized state near the conduction (valence) band. Instead, independent of the nature of the dopant, addition or removal of electrons to or from the polymer chain is invariably accommodated by formation soliton states near mid-gap. Calculations of the absorption coefficient (α) show that a soliton kink on a chain suppresses the interband transition, whereas transitions involving the soliton level are found to have a significantly enhanced absorption cross-section. The results are in agreement with the experimental absorption spectra obtained from trans-($\text{CH})_x$ lightly doped with AsF_5 .

Takayama, Lin-Liu and Maki (TLM) (7) considered the continuum limit of the linear chain model introduced by SSH to describe $(\text{CH})_x$. Their analytical results are in agreement with the numerical results of SSH and allow explicit calculation of wave functions and matrix elements. The effective Hamiltonian for the continuum model is

$$H = -iv_F \sigma_3 \frac{\partial}{\partial x} + \Delta(x) \sigma_1 \quad [9]$$

with the Fermi velocity $v_F = 2 t_0 / a$. The σ 's are Pauli matrices, a and t_0 are the lattice constant and nearest neighbor transfer integral, respectively, of the uniform (undimerized) chain, and $\Delta(x)$ is the order parameter describing the dimerization pattern. To calculate $\alpha(\omega)$, Suzuki et al (11) have first obtained the matrix elements of the momentum operator, which can be expressed in the continuum model in the form $p_x = 2M_x i \sigma_3$ where $M_x = -i \int \phi(x, y, z) \frac{\partial}{\partial x} \phi(x-a, y, z) dx dy dz$ and ϕ is the atomic $\pi(p_z)$ orbital of the carbon atoms. They find that for a perfect dimerized chain ($\Delta(x) = \Delta_0$), the interband absorption coefficient (per carbon atom), for transitions from valence band (VB) to conduction band (CB), is

$$\alpha_1(\omega) = A f_1(\omega) \quad [10]$$

$$f_1(\omega) = \left(\frac{E_g}{\hbar\omega} \right)^2 \frac{E_g}{[\hbar\omega - E_g]^2}^{1/2} \quad [11]$$

where $A = (16\pi^2 \hbar e^2 |M_x|^2) / m_e^2 n c W E_g$, $E_g = 2\Delta_0$, $W = 4t_0$, and m_e and e are the electron mass and charge, c is the velocity of light, and n is the index of refraction. Although $\alpha_1(\omega)$ diverges as $(\hbar\omega - E_g)^{-1/2}$, this square root singularity will be smeared out by disorder, interchain coupling, or fluctuations.

For a trans-(CH)_x chain with a static kink ($\Delta(x) = \Delta_0 \tanh(x/\xi)$), the soliton formation energy takes the minimum value $\frac{2}{\pi} \Delta_0$ with $\xi = \xi_0 = \hbar v_F / \Delta_0$, and one bound state appears at mid-gap. Using the results of TLM Suzuki et al calculated α_s , for transitions between a soliton level (S) and the band states to be

$$\alpha_s(\omega) = A \frac{\pi^2 \xi_0^2}{a} f_s(\omega) \quad [12]$$

$$f_s(\omega) = \frac{E_g}{[4(\hbar\omega)^2 - E_g^2]^{1/2}} \operatorname{sech}^2 \left(\frac{\pi}{2} \frac{[4(\hbar\omega)^2 - E_g^2]^{1/2}}{E_g} \right) \quad [13]$$

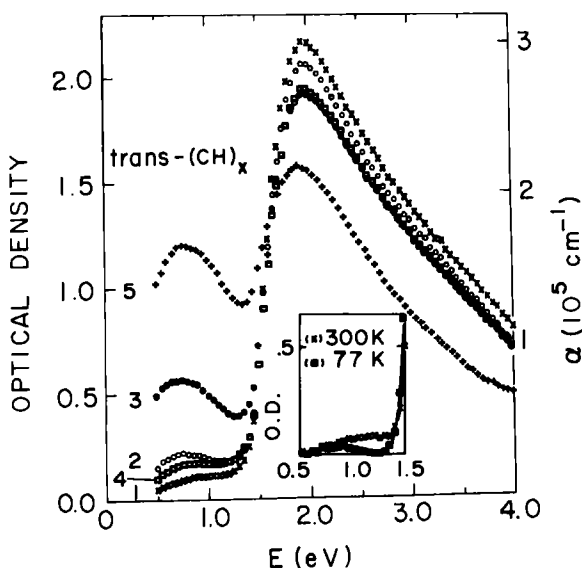
The factor $(\pi^2 \xi_0^2 / a)$ in eq. 12 indicates an enhancement of the soliton transition resulting directly from the delocalization of the soliton wave function ($\xi/a \gg 1$). For $W = 10$ eV and $E_g = 1.5$ eV, one finds $\xi_0 \sim 7a$ and $(\pi^2 \xi_0^2 / a) \sim 70$. As a result, the transitions involving the soliton level should be observable even at extremely small concentrations. The interband $\alpha'_1(\omega)$ (per carbon atom) in the presence of a kink is correspondingly decreased; i.e.,

$$\alpha'_1(\omega) = \alpha_1(\omega) \left(1 - \frac{\ln 2}{2N} \frac{\pi^2 \xi_0^2}{a} \right) \quad [14]$$

It is interesting to note that the presence of a soliton on a chain breaks the translational symmetry of the lattice. As a result, the momentum conserving (vertical) interband transition is quenched and α'_1 is due to non-vertical terms (11,18).

The optical absorption spectra of trans-(CH)_x are shown in Fig. 4.

Figure 4
Absorption spectra of
trans-(CH)_x: curve 1,
undoped; curve 2, 0.01%
AsF₅; curve 3, 0.1%
AsF₅; curve 4, compensated
by NH₃; and curve 5, 0.5%
AsF₅. The inset shows
the temperature dependence
of $\alpha(\omega)$ for undoped
trans-(CH)_x.



Curves 1 through 5 on Fig. 4 show the effects of doping; 1-undoped, 2 very lightly doped with AsF₅ (about 0.01%), 3-lightly doped with AsF₅ (about 0.1%), and 4-subsequently compensated with NH₃; all data are from the same (CH)_x film. The doping was carried out *in situ* using extreme care so that the results could be directly and quantitatively compared. Curve 5 on Fig. 4 was obtained with a separate film doped to a somewhat higher level (~.5%); quantitative comparison was possible through normalization of the absorption curve prior to doping. The effect of temperature on undoped (CH)_x is shown in the inset; data obtained at 77 K and 4.2 K are indistinguishable.

The strong absorption band with edge at 1.4 eV and peak at 1.95 eV has been attributed to the direct interband transition in a one-dimensional (1-d) band structure, and can be viewed as arising from a transition from the 1-d peak in the density of states in the valence band to that in the conduction band. The rounding appears to shift the position of the peaks in the VB and CB densities of states by about 0.2 eV. In addition to the main peak, a weak absorption (see the inset to Fig. 4) is observed centered near 0.9 eV, corresponding to a transition between the peak in VB density of states and a

level inside the gap. Relative to the VB edge, the gap state occurs at about 0.7 eV; i.e., near mid-gap.

On doping with AsF_5 , the low energy absorption grows with increasing concentration and shifts toward lower energy; compensation decreases \propto almost to the same level as before doping. The strength of the main absorption decreases on doping with AsF_5 , but it does not recover after subsequent compensation with NH_3 .

The characteristic features of the low energy and main absorption bands can be explained if we assume that the doping proceeds through formation of positive charged solitons (S^+) and that the low energy absorption band is associated with the transition from the valence band into the S^+ level to form a neutral soliton (S^0). The low energy absorption, seen in the low temperature data of inset to Fig. (4), indicates that soliton states are also present in the undoped material. As the number of S^+ increases with AsF_5 doping, the strength of the low energy transition ($S^+ + e \rightarrow S^0$) grows proportionally. Further, the calculations presented above predict that the interband transition will be suppressed with the introduction of soliton kinks, again in agreement with the experimental results. On reacting with NH_3 , the S^+ level is compensated, and the low energy band decreases accordingly. On the other hand, the interband transition does not recover to its initial strength on compensation, since the π -electron kinks remain on the chain; only the charged center is compensated.

The integrated intensity of the low energy absorption band after 0.1% doping (curve 3) is about one tenth of that of the main absorption band of the undoped sample. We infer from eq. 12 a value for $\pi^2(\xi/a) \sim 10^2$ in good agreement with the theoretical value. The uniform decrease in intensity of the interband transition is also evident in curve 3 and corresponds to about a 10% reduction relative to the undoped sample.

PHOTOEXCITATIONS IN POLYACETYLENE (13-14)

Interest in photoexcitation studies of $(\text{CH})_x$ has been stimulated by the recent calculations of Su and Schrieffer (19), who considered direct injection of an e-h pair and studied the time evolution of the system. Their principal result was the conclusion that in trans-($\text{CH})_x$ a photo-injected e-h pair evolves to a soliton-antisoliton pair in a time of order of the reciprocal of an optical phonon frequency. This is the central experimental question addressed in this section: Are solitons photogenerated in polyacetylene?

The proposed photogeneration of charged solitons (20) has clear implications which can be checked through studies of photoluminescence and photoconductivity. We therefore present and discuss in this section the experimental results obtained with these two complementary techniques in cis- and trans-($\text{CH})_x$. In the scattered light spectrum from cis-($\text{CH})_x$, we find a relatively broad luminescence structure peaking at 1.9 eV, near the interband absorption edge, together with a series of multiple order Raman lines (14,21). (see Fig 5A)

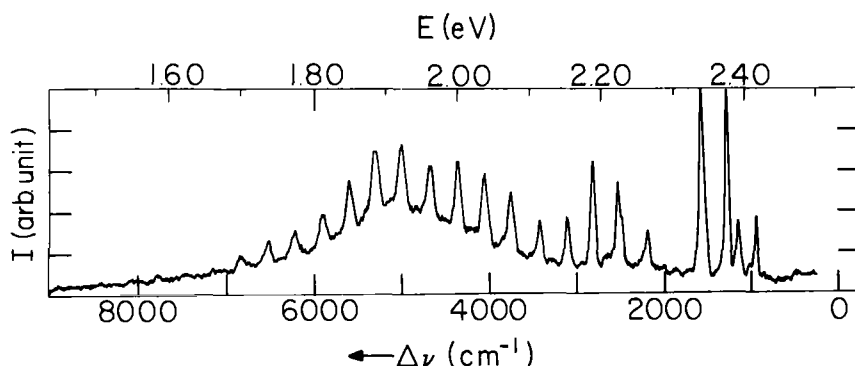


Figure 5A

Photoluminescence and multiple overtone Raman structure from $\text{cis}-(\text{CH})_x$ obtained at 7 K. The scattered light intensity is plotted as a function of frequency shift ($\Delta\nu$) away from the 2.54 eV laser excitation.

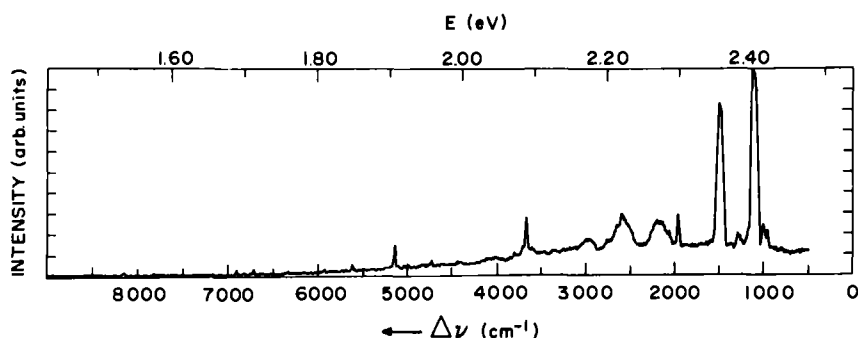


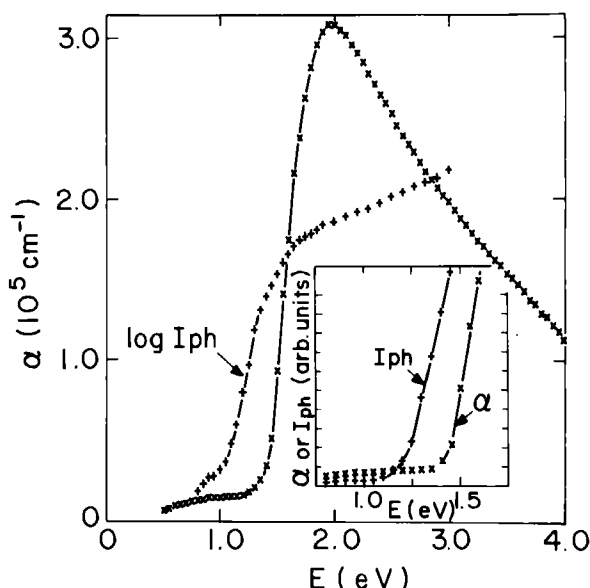
Figure 5B

Scattered light spectrum from $\text{trans}-(\text{CH})_x$ obtained at 7 K using 2.54 eV laser excitation. The intensity of any luminescence near the onset of interband absorption (1.5 eV) is less than that of the $\text{cis}-(\text{CH})_x$ peak (Fig. 5A) by at least a factor of fifty.

Through a measurement of the excitation spectrum, we have found that the luminescence turns on sharply for excitation energies greater than 2.05 eV, implying a Stokes shift of 0.15 eV. The main peak in the luminescence structure at 1.9 eV has a half-width of about 0.1 eV and a Lorentzian energy profile. The luminescence intensity is essentially independent of temperature from 7 K to room temperature. Isomerization of the same sample to $\text{trans}-(\text{CH})_x$ quenches the luminescence; we find no indication of luminescence near the interband absorption edge of $\text{trans}-(\text{CH})_x$ even at temperatures as low as 7 K. (Fig. 5B) (14).

The photocurrent (I_{ph}) spectrum for trans-(CH)_x is plotted in Figure 6

Figure 6
Comparison of the photocurrent ($\log I_{ph}$) and optical absorption coefficient (α) as a function of photon energy in trans-(CH)_x . Note the threshold of I_{ph} near 1 eV, well below the onset of interband absorption. I_{ph} and α are compared on a linear scale in the inset.



and compared with the energy dependence of the absorption coefficient for trans-(CH)_x , $\alpha(\omega)$ (13). I_{ph} has a threshold at 1.0 eV, well below the interband absorption edge at 1.5 eV, implying the presence of states deep inside the gap. This is clearly shown in the inset to Fig. 6 where the photoresponse and absorption are compared on a linear scale. The threshold for the generation of free carriers is better seen in the logarithmic plot of I_{ph} versus photon energy (Fig. 6). The absence of structure in $\alpha(\omega)$ at the onset of I_{ph} implies a low quantum efficiency at threshold. The free carrier generation efficiency rises exponentially above the threshold and changes to a slow increase above the onset of the interband transition.

Measurements of I_{ph} in cis-(CH)_x were also attempted. Under similar conditions any photoresponse in samples of 80% cis-rich (CH)_x was below the noise level of our experiment. The upper limit on I_{ph} in cis-(CH)_x is more than three orders of magnitude smaller than I_{ph} in trans-(CH)_x . In situ isomerization of the same film resulted in the sizeable I_{ph} shown in Fig. 6.

The observation of luminescence in cis-(CH)_x but not in trans-(CH)_x , and the observation of photoconductivity in trans-(CH)_x but not in cis-(CH)_x provide confirmation of the proposal that solitons are the photogenerated carriers. In trans-(CH)_x , the degenerate ground state leads to free soliton excitations, absence of the band edge luminescence and photoconductivity. In cis-(CH)_x the non-degenerate ground state leads to confinement of the photogenerated carriers, absence of photoconductivity, and to the observed recombination luminescence.

In traditional semiconductors, photoconductivity and recombination luminescence are intimately related, and both are observed after photoexcitation. Photoconductivity indicates the presence of free carriers generated by the absorbed photons. Although the subsequent recombination of these photogenerated carriers can take place either radiatively or non-radiatively, recombination luminescence is commonly observed, at least at low temperatures. The fundamental differences between such traditional data and those obtained from polyacetylene can be seen by comparison of the $(\text{CH})_x$ results with results obtained from cadmium sulfide (CdS). The luminescence and multiple order Raman scattering data from CdS (22) are similar to the results obtained from $\text{cis}-(\text{CH})_x$. However, in CdS , a strong photoconductive response is observed for photon energies just above the band edge, (23) whereas in $\text{cis}-(\text{CH})_x$ significant photogeneration of free carriers is not observed even for photon energies 1 eV above the band edge. Isomerization to $\text{trans}-(\text{CH})_x$ quenches the luminescence at all temperatures, but turns on the photoconductivity. In neither isomer is the traditional combination of effects observed. These unique experimental results, therefore, lead us to consider the proposed photogeneration of solitons in more detail.

A schematic diagram of the photogeneration of a charged soliton-antisoliton pair in $\text{trans}-(\text{CH})_x$ is shown in Fig. 7.

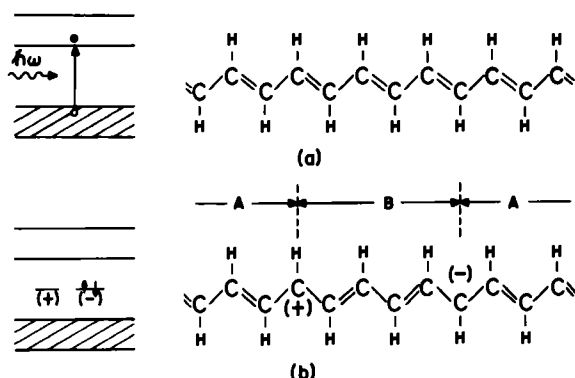


Figure 7

- Band diagram and chemical structure diagram for $\text{trans}-(\text{CH})_x$. The band diagram shows schematically the absorption of a photon and the creation of an e-h pair.
- A $\text{trans}-(\text{CH})_x$ chain containing a charged soliton-antisoliton pair. Since regions A and B are degenerate, the solitons are free and can move apart with no cost in energy. The corresponding band diagram shows the mid-gap states associated with the two solitons; one empty (+) and the other doubly occupied (-). The soliton deformations are shown as localized on a single site, whereas the deformation should be spread over about 15 lattice constants.

The incident photon (for $\hbar\omega \gg 2\Delta$) generates an e-h pair within the rigid lattice (Fig. 7a). The system rapidly evolves to a charged soliton pair (Fig. 7b) as shown by Su and Schrieffer.⁽¹⁹⁾ After a time of order 10^{-13} sec, their results imply the presence of two kinks separating degenerate regions. Because of the precise degeneracy of the A and B phases, the two charged solitons are free to move in an applied electric field and contribute to the photoconductivity. The simultaneous absence of band edge luminescence in trans-(CH)_x even at the lowest temperatures is consistent with the proposed photogeneration of charged soliton-antisoliton pairs. In this case, band edge luminescence cannot occur since there are no electrons and holes. The charged carrier pair consists of two kinks with associated mid-gap states which cannot give rise to band edge luminescence.

The topological degeneracy in trans-(CH)_x is not present in cis-(CH)_x, so that soliton photogeneration would not lead to photoconductivity in the cis-isomer. Since the cis-transoid configuration (Fig. 8a) has a lower

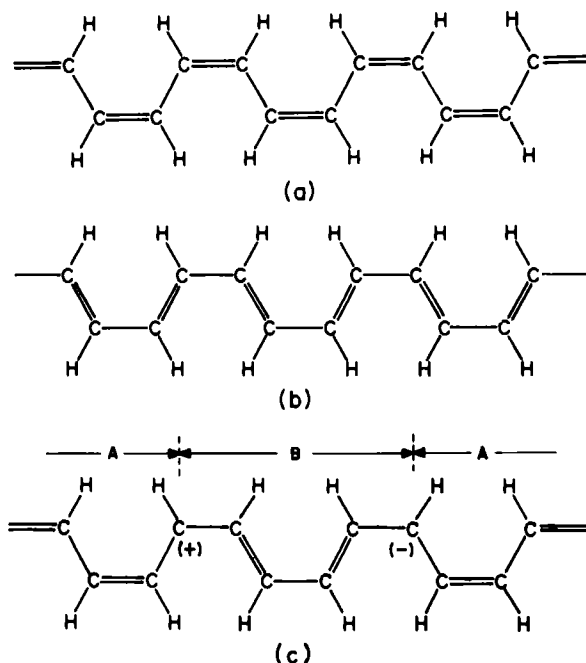


Figure 8

- a. cis-(CH)_x; cis-transoid configuration (lower energy).
- b. cis-(CH)_x; trans-cisoid configuration (higher energy).
- c. A cis-(CH)_x chain containing a charged soliton-antisoliton pair. Since the regions A and B are not degenerate, the solitons are confined; the farther apart the greater the energy. The soliton deformations are shown as localized on a single site, whereas the deformation should be spread over many lattice sites.

energy than the trans-cisoid configuration (Fig. 8b), domain walls would separate non-degenerate regions (Fig. 8c). Although neither of the limiting forms of Fig. 8a and 8b is literally correct, we may think of the ground state as essentially equivalent to Fig. 8a, with the structure of Fig. 8b being slightly higher in energy. Consider then the photoexcitation of a charged soliton-antisoliton pair as shown schematically in Fig. 8c. The energy required is

$$E_{\text{tot}} = 2E_s + n \Delta E_0 \quad [15]$$

where E_s is the energy for creation of a single soliton (analogous to the soliton creation energy in trans-(CH)_x, n is the number of CH monomers separating the two kinks, and ΔE_0 is the energy difference between the two configurations (Fig. 8a and 8b), expected to be a small fraction of the full gap $2\Delta \sim 2.0$ eV in cis-(CH)_x. Thus, as they begin to form, the two solitons would be "confined", or bound into a polaron-like entity; the farther apart, the greater the energy. As a result, one expects the photogenerated pair to quickly recombine, thereby, quenching the photoconductivity in cis-(CH)_x.

The luminescence results in cis-(CH)_x as described earlier provide some insight into the dynamics of system. The energy given off to phonons during the formation of the lattice distortion would lead to a Stokes shift of the luminescence relative to the minimum energy needed to make a free e-h pair. As indicated above, a Stokes shift of $\Delta E_s = 0.15$ eV is observed relative to the onset of photoexcitation at 2.05 eV. The magnitude of ΔE_s is, however, much less than that expected for the formation of either two free solitons or two free polarons. In the case of two free solitons, the corresponding electronic levels would be near mid-gap with the luminescence energy going to zero in the limit of $(\Delta E_0/2\Delta) \rightarrow 0$. In this limit for a well-separated electron-polaron and hole-polaron, (24,25) the corresponding luminescence energy would be $\hbar\omega_L = 2\Delta/\sqrt{2} = 1.4\Delta$ with an implied Stokes shift of $\Delta E_s = .6\Delta$. The much smaller experimental value of $\Delta E_s \sim .07\Delta$, therefore, indicates that the absence of a degenerate ground state in cis-(CH)_x has made qualitative changes in the dynamics of photoexcitations in the two isomers.

The suggestion that confinement of photogenerated carriers plays an important role in cis-(CH)_x may provide an explanation for the dramatic difference in the light scattering spectra of the two isomers as shown in Fig. 5. Brazovskii and Kirova have pointed out that in general the total gap in a commensurate Peierls distorted system arises from a combination of two terms (24); i.e.,

$$\Delta(x) = \Delta_e + \Delta_i(x) \quad [16]$$

where $\Delta_i(x)$ is the intrinsic Peierls gap stabilized by π -electrons, and Δ_e is an extrinsic contribution which is incompressible and arises from σ -electron dimerization. In trans-(CH)_x $\Delta_e = 0$ and they suggest a value of $\Delta_e = .3$ eV for cis-(CH)_x. This is in reasonable agreement with the larger optical gap in cis-(CH)_x; i.e., $2\Delta_{\text{cis}} \sim 2.0$ eV compared to $2\Delta_{\text{trans}} \sim 1.5$ eV. A sizeable Δ_e together with a relatively small Stokes shift indicate that because of the confinement the distortion of cis-(CH)_x lattice in the presence of an electron-hole pair may be in the shallow polaron limit described by Brazovskii and Kirova (26). This is the opposite limit to the case of trans-(CH)_x lattice where large amplitude distortions in $\Delta(x)$ are precursor to the formation of a soliton-antisoliton pair upon photoinjection of an electron-hole pair (19). Since the scattered light spectra in principle reflect the response of the lattice to the photoexcited carriers, the dramatic difference between the experimental results from cis- and trans-(CH)_x may indeed be due to the confinement.

The spectral dependence of the photoconductivity in trans-(CH)_x provides supportive evidence for the proposed photogeneration of solitons. The minimum energy required for photogeneration of a positive and negative soliton pair is $2(2\Delta/\pi) < 2\Delta$. (7,24) Since the direct photogeneration of a soliton pair requires a significant lattice distortion simultaneous with the electronic transition, the quantum efficiency near threshold is expected to be small and to increase exponentially as $\hbar\omega$ approaches 2Δ , in a manner similar to an Urbach tail (27). For $\hbar\omega > 2\Delta$, the interband transition has a quantum efficiency of order unity for direct e-h pair production so that the overall soliton-antisoliton photogeneration process should be only a weak function of energy. The experimental results are in agreement; I_{ph} rises rapidly and exponentially above 1 eV $2(2\Delta/\pi)$ and more slowly above 1.5 eV $\approx 2\Delta$. Thus the spectral dependence of I_{ph} for trans-(CH)_x is consistent with photogeneration of charged soliton-antisoliton pairs. The knee in I_{ph} at $\hbar\omega \sim 2\Delta$, the onset of interband absorption, must be contrasted with the sharp rise typically seen at that point in traditional semiconductors where electrons and/or holes are the photoinjected carriers.

The broadening of the absorption edge can be expected from static disorder, or from a combination of quantum and thermal fluctuations in $\Delta(x)$, the local energy gap parameter (8,24). Such dynamical fluctuations generate random dynamical distortions in the crystal. Since the electronic states that contribute to the absorption edge are determined in accord with the Frank-Condon principle by the instantaneous ($\sim 10^{-16}$ sec) configuration of the lattice, these fluctuations lead to a distribution of short lived states ($\sim 10^{-18}$ sec) below the band edge. As a result, the absorption edge is smeared and an Urbach tail is generated. For temperatures less than a typical phonon energy ($\hbar\omega_p \sim 1500 \text{ cm}^{-1}$) Brazovskii has shown that the energy dependence of such an Urbach absorption tail varies as (8,24)

$$\alpha(\omega) \sim \exp\left(-\frac{\Delta}{\hbar\omega_p} \left(2 - \frac{\hbar\omega}{\Delta}\right)^{3/2}\right) \quad 17$$

for $\hbar\omega < 2\Delta$. In this case, the associated microfield (27) arises from the quantum fluctuations instead of disorder or optical phonons as is the case for ionic or covalent semiconductors. A fit of the exponential rise of the photocurrent below the band edge (see Fig. 6) to equation 16 results in $\hbar\omega_p \sim 0.2 \text{ eV}$, in agreement with the lattice dynamic results for trans-(CH)_x (12,16).

CONCLUSION

The experimental results from polyacetylene reviewed in this paper show that a number of unusual effects are observed in the lattice dynamics (monitored by IR and Raman spectroscopy), the band structure (seen through optical absorption), and the photoexcitation dynamics (monitored by photoluminescence and photoconductivity). All of these results have been coherently interpreted in terms of the soliton model. The theoretical interpretation of the experimental results has been based on a simple 1-d Hamiltonian, which describes the Peierls distortion in a system of noninteracting electrons with strong coupling to the lattice. Solitons are the natural nonlinear excitations of such a system. It is shown that, unlike the traditional semiconductors, where electrons and holes can be stable excitations, in polyacetylene the lattice is inherently unstable to their presence. Injection of an electron-electron (hole-hole) pair by chemical doping or photoinjection of an electron-hole pair distort the lattice leading to formation of a soliton-antisoliton pair in a time scale of $\sim 10^{-13}$ seconds; thus preventing the system from relaxing by any other means

EFFECTS IN LATTICE DYNAMICS: The local distortion generated by the presence of a soliton on a chain gives rise to a set of infrared active

vibrational modes which are due to the large amplitude oscillations of charge across the soliton center. The agreement between the model calculations and the experimentally observed IAVM reflects the specific odd-function character of the soliton wave function. We have been able to achieve a quantitative understanding of the Raman frequencies and the soliton induced infrared absorption in $\text{trans}-(\text{CH})_x$ and $\text{trans}-(\text{CD})_x$. The success of the theory implies the basic validity of the one-electron approach, which Coulomb interactions playing only a relatively minor role. The detailed agreement between the calculated and experimentally observed infrared active vibrational modes provides specific experimental evidence of soliton doping in polyacetylene.

EFFECTS SEEN IN THE BANDSTRUCTURE: Explicit evidence for generation of soliton mid-gap states, from a joint theoretical and experimental study of the effects of dilute doping on the optical absorption spectra of $\text{trans}-(\text{CH})_x$, have been reviewed. Calculations within the soliton model has lead to three specific predictions: (1) A soliton absorption band, independent of the nature of the dopant, should appear near the middle of the absorption gap. (2) The intensity of transitions involving the soliton level should be enhanced by about two orders of magnitude. (3) The existence of a soliton on a chain should suppress uniformly the entire interband transition. All of these effects are observed experimentally with magnitudes in agreement with the theoretical results.

EFFECTS SEEN IN PHOTOEXCITATIONS: The combined results of two complementary experiments; i.e., photoluminescence and photoconductivity in $\text{cis}-(\text{CH})_x$ and $\text{trans}-(\text{CH})_x$ indicate that the photoinjected carriers are indeed the soliton-antisoliton pairs. The principle results are summarized in the following table:

	<u>Cis</u> -(CH) _x	<u>Trans</u> -(CH) _x
Ground State	Non-degenerate	Degenerate by symmetry
Nonlinear excitation	Confinement	Stable soliton-antisoliton pair
Band-edge luminescence	Yes	No
Photoconductivity	No	Yes

The experimental results presented thus confirm the theoretical expectation that the $\text{trans}-(\text{CH})_x$ lattice is unstable to the presence of any pair of electron or hole excitations. The strong coupling of the electronic structure to the lattice provides a naturally fast route for the lattice to distort and transform the simple charge excitation pairs into soliton-antisoliton pairs. As a result of these unusual features, polyacetylene is fundamentally different from conventional semiconductors.

The simple one-dimensional structure of trans -polyacetylene provides the means for experimental realization of a variety of mathematical problems. For example, the concept of soliton as a nonlinear solution of the Dirac equation has been of considerable mathematical interest. The recent mapping (25,28) of SSH Hamiltonian onto the two dimensional Dirac equation has shown a unifying base to be shared by condensed matter and particle physicists. The possibility of experimental studies of solitons in polyacetylene is, therefore, of broad interest to many fields of physics.

Acknowledgement: This work is a review of the result of fruitful collaborations with T. -C. Chung, L. Lauchlan, A. G. MacDiarmid, M. Ozaki, A. Pron, and N. Suzuki at the University of Pennsylvania, and E. G. Mele and M. J. Rice of Xerox. The review was prepared under support from the Army Research Office (DAAG29-81-K-0058).

References:

1. For a review of the initial works on polyacetylene see: Heeger, A. J. and MacDiarmid, A. G., The Physics and Chemistry of Low Dimensional Solids Ed. by Louis Alcacer (D. Reidel Publishing Co., 1980) p. 353.
2. Highly Conducting One-Dimensional Solids, Ed. by J. T. Devreese, R. P. Evrard, and V. E. van Doren (Plenum Press, N. Y. 1978) contains several reviews on this topic.
3. Rice, M. J., and Strassler, S., Solid State Commun. **13**, 125 (1973).
4. Baughman, R. H., Hsu, S. L., Pez, G. P., and Signorelli, A. J., J. Chem. Phys., **68**, 5405 (1978).
5. Su, W. P., Schrieffer, J. R. and Heeger, A. J., Phys. Rev. Lett. **42**, 1698 (1979); ibid. Phys. Rev. B **22**, 2099 (1980).
6. Rice, M. J., Phys. Lett. **71A**, 152 (1979).
7. Takayama, H., Lin-Liu, Y. R. and Maki, K., Phys. Rev. B **21**, 2388 (1980).
8. Brazovskii, S., JETP Lett. **28**, 656 (1978); ibid. JETP **78**, 677 (1980).
9. Nechtschein, M., Devreux, F., Greene, R. L., Clarke, T. C. and Street, G. B., Phys. Rev. Lett. **44**, 356 (1980).
10. Weinberger, B. R., Ehrenfreund, E., Pron, A., Heeger, A. J. and MacDiarmid, A. G., J. Chem. Phys. **72**, 4749 (1980).
11. Suzuki, N., Ozaki, M., Etemad, S., Heeger, A. J. and MacDiarmid, A. G., Phys. Rev. Lett. **45**, 1209 (1980); Erratum Phys. Rev. Lett. **45**, 1483, (1980).
12. Etemad, S., Pron, A., Heeger, A. J., MacDiarmid, A. G., Mele, E. G., and Rice, M. J., Phys. Rev. **B23**, 5137 (1981).
13. Etemad, S., Mitani, M., Ozaki, M., Chung, T. -C., Heeger, A. J., and MacDiarmid, A. G. Solid State Communications (in press).
14. Lauchlan, L., Etemad, S., Chung, T. -C., Heeger, A. J., and MacDiarmid, A. G., Phys. Rev. B (in press).
15. Fincher, C. R., Jr., Ozaki, M., Heeger, A. J., and MacDiarmid, A. G., Phys. Rev. **B19**, 4140 (1979).
16. Mele, E. J. and Rice, M. J., Solid State Commun **34**, 339 (1980).
17. Mele, E. J., and Rice, M. J., Phys. Rev. Lett. **45**, 926 (1980).

18. Gamal, J. T. and Krumhansel, J. A. (preprint), (1981).
19. Su, W. P. and Schrieffer, J. R., Proc. Nat. Acad. of Sci. 77, 5626 (1980).
20. Etemad, S., Ozaki, M., Heeger, A. J. and MacDiarmid, A. G., Proc. of the "International Conf. on Low Dimensional Synthetic Metal," Helsingor, Denmark, August 1980, *Chemica Scripta* (in press).
21. Lichtmann, L. S., Sarhangi, A. and Fitchen, D. C., Solid State Commun. 36, 869 (1980).
22. Leite, R. C. C., Scott, J. F., and Damen, T. C., Phys. Rev. Lett. 22, 780 (1969); Klein, M. V. and Parto, S. P. S., Phys. Rev. Lett. 22, 782 (1969).
23. See, for example, Bube, R. H., Photoconductivity of Solids, Wiley and Sons, N. Y. 1960, p. 230 and 391.
24. Brazovskii, S. and Kirova, N., Pisma Zh. Eksp. Teor. Fiz. 33, 6 (1981).
25. Campbell, D. and Bishop, A., Phys. Rev. B 24, 48-59 (1981).
26. Using the results of Ref. 24, the experimental value of the Stokes shift implies a shallow distortion, with a maximum value of $\sim .15\Delta$ and full width of $\sim 40a$, in the gap parameter $\Delta(x)$.
27. See Dow, J. D. and Redfield, D., Phys. Rev. B 5, 594 (1972) and references therein.
28. Jackiw, R. and Schrieffer, J. R., Phys. B 190 [FS3], 253 (1981).

This Page Intentionally Left Blank

QUASI SOLITONS: A CASE STUDY OF THE DOUBLE SINE-GORDON EQUATIONS

P. Kumar and R.R. Holland

Physics Department
University of Florida
Gainesville, Florida 32611

We report here results from the numerical simulation of the scattering between a soliton-antisoliton pair for the generalized double sine-Gordon equation.

INTRODUCTION

The double sine-Gordon equation (DSGE) is a curious non-linear partial differential equation in that it admits two distinct classes of soliton solutions.⁽¹⁻³⁾ Furthermore, unlike an ideal soliton equation, where the solutions undergo elastic scattering, the scattering between DSGE solitons is weakly inelastic.⁽³⁾ During a collision, small amplitude waves are emitted that carry away the energy. The consequences of the weak inelasticity are two-fold. On the one hand, the kinematics of the collision process is changed. On the other hand, perturbation theory methods have been developed by Kaup and Newell⁴ and by Scott and McLaughlin⁵ and applied to cases where sine-Gordon equation is weakly perturbed. If the DSGE can be expressed as a perturbation on the sine Gordon equation, the present numerical simulation can provide a test of the analytic approaches.

The DSGE is a class of non-linear partial differential equations described by

$$\phi_{tt} - c^2 \phi_{xx} = \frac{\omega_0}{1-\alpha^2} \sin\phi(\cos\phi + \alpha) \quad (1)$$

Here the subscripts denote a derivative and the parameter α describes different members of the class. For $\alpha=0$, we recover the sine-Gordon equation. The equation appears in several physical contexts. In superfluid ^3He ,⁽¹⁻²⁾ as well as in a special case of self-induced transparency (SIT) with level degeneracy,⁽²⁾ α takes the value $1/4$ and has been analyzed previously. A different version of Eq. (1), with RHS carrying a minus sign also occurs frequently in different physical contexts e.g. poling process in PVF_2 ,⁽⁶⁾ and the more common equation for SIT with level degeneracy.⁽⁷⁾ However this equation does not carry the two distinct classes of solitons, it is more like a perturbation on SG equation and has been studied analytically by Bullough et.al.⁽⁷⁾

Eq.(1) may be derived from a potential $V(\phi) = \frac{\omega_0}{2(1-\alpha^2)} (\cos\phi + \alpha)^2$, , drawn schematically in Fig. (1). It consists of valleys of minima

at $\phi = \phi_0 = \cos^{-1}(-\alpha)$ and at $\phi = 2\pi - \phi_0$ separated by maxima at $n\pi$; $n=0,1,2,\dots$. The prefactor $1/(1-\alpha^2)$ ensures that the potential has curvature ω_0^2 at the minima. The two solitons correspond to the system being at adjacent minima asymptotically ($x \rightarrow -\infty$ and $+\infty$) and connected via the potential maxima. The type II soliton $(-\phi_0, \phi_0)$ is connected via the

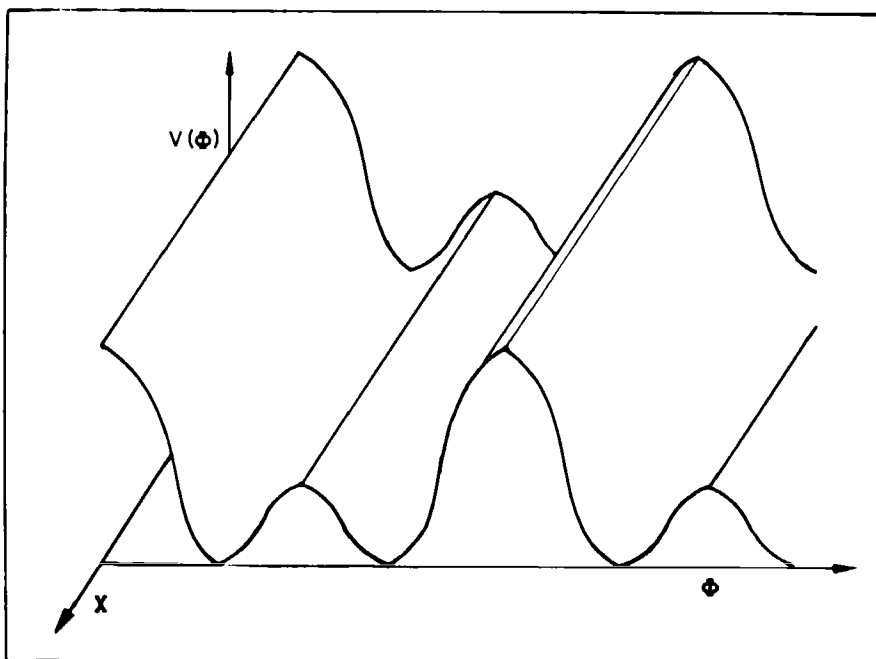


Fig. 1. The potential $V(\phi)$ as a function of ϕ and χ . The minima are at $\phi_0 = \cos^{-1}(-\alpha)$

higher peak at $\phi=0$ and is more massive than the type I soliton $(\phi_0, 2\pi - \phi_0)$ which is connected via the lower peak at $\phi=\pi$. There are topological implications in that a type I soliton must be followed either by a type II soliton or a type I antisoliton ($\phi_0 < 0$). Two type I solitons have infinite repulsion. Subject to the boundary conditions mentioned above, the two solutions are

$$\frac{\tan \phi/2}{\tan \phi_0/2} = \coth \left[(x-ut)/\zeta \sqrt{1-v^2} \right] \quad \text{type I} \quad (2)$$

$$= \tanh \left[(x-ut)/\zeta \sqrt{1-v^2} \right] \quad \text{type II} \quad (3)$$

where $\zeta = c/\omega_0$ is a characteristic length in the problem and $u=vc$ is the velocity of the soliton. The energy of these solutions is given by

$$E_I = \frac{2Ac\omega_0}{\sqrt{1-\alpha^2}} (\sin \phi_0 + \alpha(\phi_0 - \pi)) \quad (4)$$

$$E_{II} = \frac{2Ac\omega_0}{\sqrt{1-\alpha^2}} (\sin \phi_0 + \alpha \phi_0) \quad (5)$$

where A is an overall energy scale factor, clearly $E_{I2} < E_{II}$. The limit $\alpha=1$ is singular. If we drop the prefactor $(1-\alpha^2)$ in the potential, it becomes a $\cos \phi/2$ potential. The solitons then are $(-\pi, \pi)$ solutions and are given by

$$\tan \phi/2 = \frac{x-ut}{\zeta \sqrt{1-v^2}} \quad (6)$$

$$E = 2\pi A \omega_0^2 \zeta \quad (7)$$

Linear stability analysis shows that these solitons are stable. The normal mode spectrum in each case includes a bound state at $\omega=0$, corresponding to the translational mode and a continuum of wave like excitations whose frequency is given by the dispersion relation $\omega^2 = \omega_0^2 + c^2 k^2$. The phonons are not reflectionless anymore. In a collision between a soliton-antisoliton ($S\bar{S}$) pair, many phonons are emitted that take the energy away and make the collision inelastic.

NUMERICAL SIMULATION RESULTS

Computer simulation of soliton dynamics has historically been an important and essential source of information. Beginning with the works of Fermi Pasta and Ulam, numerical simulation has been an integral part of analyses of soliton dynamics. These have been extensively reviewed by Bullough⁽⁷⁾ and by Makhankov.⁽⁸⁾ The numerical study of DSG (for $\alpha=1/4$) has been previously reported by Maki and Kumar,⁽¹⁾ Kitchenside et.al.⁽²⁾ and by Shieffman and Kumar.⁽³⁾ We study here the α dependence of the properties of DSG solitons.

As before,⁽³⁾ the numerical calculations were done using discrete space-time steps. At the initial time, the soliton-antisoliton pair was described by Eqs.(2) or (3). The solitons were launched towards each other from positions sufficiently far apart (5ζ) so that the effect of interaction was negligible. At any time, the data set consisted of two vectors $\phi(x_i, t)$ and $\phi(x_i, t-\Delta t)$ which was used to generate $\phi(x_i, t+\Delta t)$ using simple trapezoidal rule integrations. The profile vectors were shifted in time and $\phi(x_i, t-\Delta t)$ was dropped. In case of $S\bar{S}$ bound states, $\phi(x=0)$ was stored separately at all times. Such a procedure applied directly, caused wild fluctuations in $\phi(x_i, t+\Delta t)$. These were errors due to the discrete lattice, amplified by the non-linear terms in the partial differential equation. This instability could be controlled by averaging nearest neighbor values on the space lattice at each time. The remaining discrete lattice errors were particularly bothersome for small α where damping effects were large. For $\alpha=1$, we chose $\Delta x = .025\zeta$ and Δt was 5% smaller. For larger α , $\Delta x = .1\zeta$ (and Δt , 5% smaller) was found sufficient.

For a type I $S\bar{S}$ pair, the results of this simulation are described in Figs. (2-5). In Fig. (2), we show the classical version of the scattering matrix for $\alpha=.1$. For small incoming velocities of the $S\bar{S}$ pair, the final state is a bound state where the pair oscillates with a characteristic frequency. In a simple description of soliton as a particle moving in a potential wall, this frequency should monotonically decrease with the soliton energy. There is considerable structure in the energy dependence of the oscillation frequency. It does go to zero at a

characteristic velocity $V_C^I(\text{energy})$ that is a measure of the energy loss on collision. For velocities larger than V_C^I , the $\bar{S}\bar{S}$ pair undergoes a hard core repulsive collision. For velocity higher than another critical velocity V_C^{II} , the $\bar{S}\bar{S}$ pair is converted into a type II pair.

These features are similar to ones reported earlier for $\alpha=.25$, with the exception of the energy dependence of the oscillation frequency. The extra structure in the frequency was missed earlier in the widely separated velocity values studied. In Fig. (5), we show the two critical velocities as a function of α . For $\alpha=0$, V_C^I and V_C^{II} are both equal to zero. They increase rapidly with α but for small they remain equal. An ideal sine-Gordon $\bar{S}\bar{S}$ pair transmits through each other and, it corresponds to particle conversion

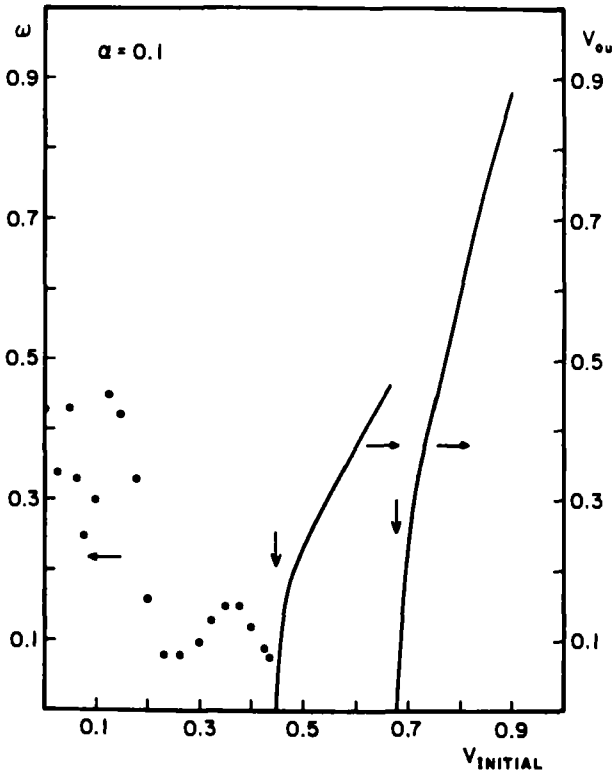


Fig. 2 The scattering matrix for $\alpha=.1$. The left hand scale refers to the frequency while the right hand scale refers to the velocity of outgoing $\bar{S}\bar{S}$ pairs. The vertical arrows indicate the thresholds for incoming velocities.

in the $\alpha \neq 0$ case. The hard-core repulsion appears only for $\alpha > .07$ at which time the V_C^I and V_C^{II} curves separate. V_C^I reaches a maximum for $\alpha=.2$ and then decreases for higher α .

Energy conservation requires for the α dependence of V_C^{II} ,

$$V_C^{II} = \left[1 - \frac{\sin \phi_0 + \alpha(\phi_0 - \pi)}{\sin \phi_0 + \alpha \phi_0} \right]^{1/2} \quad (8)$$

The observed values of V_C^{II} are slightly different from Eq.(8) due to the inelastic effects. These V_C^{II} rapidly approach one so that

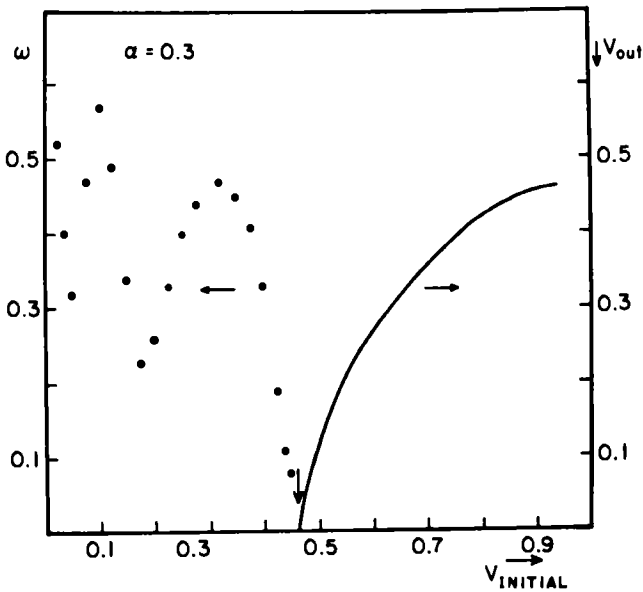


Fig. 3. The scattering matrix for $\alpha = 0.3$. The left hand scale refers to the frequency while the right hand scale refers to the velocity of outgoing SS pairs. The vertical arrows indicate the thresholds for incoming velocities.

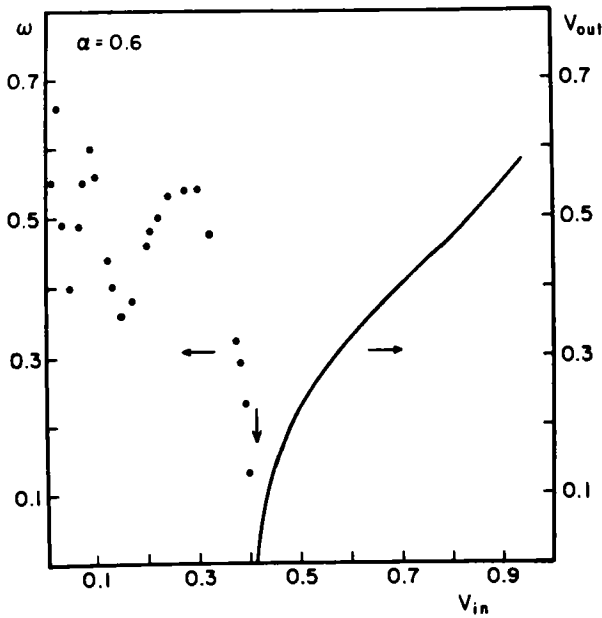


Fig. 4. The scattering matrix for $\alpha = 0.6$. The left hand scale refers to the frequency while the right hand scale refers to the velocity of outgoing SS pairs. The vertical arrows indicate the thresholds for incoming velocities.

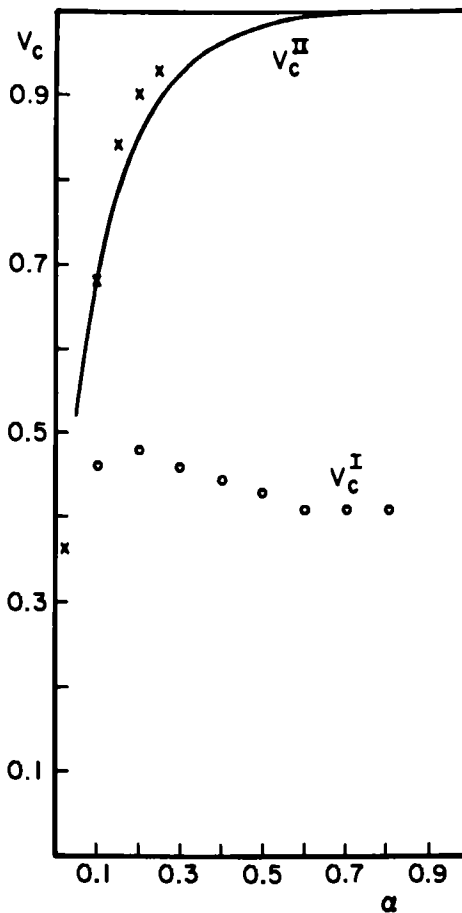


Fig. 5. The critical velocities as a function of α . The solid line shows V_c^{II} if the collisions were energy conserving. (x) are the calculated V_c^{II} while (o) are calculated V_c^I .

for $\alpha > 0.3$, the range of velocities for particle conversion becomes negligibly small. The final velocity after particle conversion, for an energy conserving collision is given by

$$V_{II} = \frac{v_I^2 - v_c^2}{1 - v_c^2} \quad (9)$$

which describes the data surprisingly well provided that we use the observed value of V_c (instead of Eq. (8)).

The effects of damping are more important at small α , as described earlier. Yet another measure of inelasticity is the rate of decay of the bound state amplitude. It decays in time with a power law as t^{-n} . The exponent n varies from $n=1.5$ for $\alpha=.1$ to $n=.2$ for $\alpha=.9$. The characteristic time scale also increases with α . In all cases however, the frequency of the bound state remains well defined. The damping remains a weak effect and the solitons can be described as quasi-solitons.

We thank A.C. Scott for encouraging discussions. This work was supported by NSF grant no. DMR 8006311 and by the Research Corporation.

REFERENCES

- (1) Maki, K. and Kumar, P., Phys. Rev. B14 (1976) 3920.
- (2) Kitchenside, P.W., Bullough, R.K. and Candrey, P.J., Creation of Spin Waves in ^3HeB , in: Bishop, A. and Schneider, T. (eds), Solitons and Condensed Matter Physics (Springer-Verlag, Berlin 1978).
- (3) Shiefman, J. and Kumar, P., Physica Scripta 20 (1979) 435.
- (4) Kaup, D.J. and Newell, A.C., Phys. Rev. B18 (1978) 5162.
- (5) McLaughlin, D.W. and Scott, A.C., Phys. Rev. A 18 (1978) 1652.
- (6) Hopfinger, A.J. Lewanski, A.J., Sluckin, T.J. and Taylor, P.L., Solitary Wave Propagation as a Model for poling in PvF_2 , in: Bishop and Schneider (red.) Solitons and Condensed Matter Physics (Springer-Verlag, Berlin 1978).
- (7) Bullough, R.K., Solitons in: Interaction of Radiation with Condensed Matter Vol. I, IAEA, Vienna 1977.
- (8) Makhankov, V., Comp. Phys. Comm. 21 (1980) 1.

This Page Intentionally Left Blank

CLASSICAL FIELD THEORY WITH Z(3) SYMMETRY

Herbert M. Ruck

Nuclear Science Division
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

ABSTRACT

We present solutions and some of their properties of a classical vector field model in two-dimensional Minkowski space with internal symmetry Z(3)—the cyclic group of order three.

1. INTRODUCTION

Cyclic symmetric field theories are of interest in solid state physics where the symmetry follows from the lattice structure of the crystals^{1,2} and in elementary particle physics where the color theory of quarks and gluons imposes cyclic transformations of the quantum numbers³.

In this short communication we are dealing with the mathematical structure of a system of nonlinear differential equations that is interesting in itself. Some physical applications are given elsewhere⁴.

2. THE Z(3) MODEL

A. The Lagrangian density of the model⁵ is a functional of a two-component vector field $\vec{\phi} = [\phi_1(t,x), \phi_2(t,x)] \in R_2$ in one time, one space dimensions $x_\mu = (t,x) \in R_2$, $\mu, \nu = 0,1$:

$$L = \frac{1}{2}(\partial_\mu \phi_1)^2 + \frac{1}{2}(\partial_\mu \phi_2)^2 - V(\phi_1, \phi_2) \quad (1)$$

$V(\phi_1, \phi_2)$ is the potential part of the Lagrangian that contains the polynomial self-interaction of the scalar fields:

$$V(\phi_1, \phi_2) = \lambda(\phi_1^2 + \phi_2^2)^2 - \nu(\phi_1^3 - 3\phi_1\phi_2^2) - \mu(\phi_1^2 + \phi_2^2) - \gamma \quad (2)$$

The metric is $g_{\mu\nu} = \text{diag}(+1, -1)$, $\mu, \nu = 0,1$; $\partial_\mu = \partial/\partial x_\mu$, $\partial_\mu^2 = \partial_\mu^2 - \partial_\mu^2$. The parameters λ, μ are positive $\lambda > 0$, $\mu > 0$. The constant γ is determined such that the potential (2) is positive definite $V \geq 0$. The parameter ν is chosen to be positive $\nu > 0$.

The group transformations of the fields that leave the Lagrangian density (1) invariant are (i) a reflection of ϕ_2 :

$$\phi_1 \rightarrow \phi_1, \quad \phi_2 \rightarrow -\phi_2 \quad (3)$$

and (ii) a Z(3) discrete rotation by 120° of the fields:

$$\phi_1 \rightarrow \phi_1 \cos(2\pi n/3) - \phi_2 \sin(2\pi n/3) , \quad (4a)$$

$$\phi_2 \rightarrow \phi_1 \sin(2\pi n/3) + \phi_2 \cos(2\pi n/3) , \quad (4b)$$

with $n = 0, 1, 2$.

The discrete symmetry $Z(3)$ Eq. (4) is imposed by the cubic term in the potential (2). When v is set zero the Lagrangian obviously has full $SO(2)$ rotational symmetry.

There exist three more representations of the Lagrangian (1): in terms of complex fields, in polar coordinates, and in terms of a matrix field.

B. With the complex fields $\phi = \phi_1 + i\phi_2$, $\phi^* = \phi_1 - i\phi_2$ the Lagrangian becomes:

$$L = \frac{1}{2} \partial_\mu \phi \partial_\mu \phi^* - \left[\lambda |\phi|^4 - \frac{1}{2} v (\phi^3 + \phi^{*3}) - \mu |\phi|^2 - \gamma \right] . \quad (5)$$

The reflection symmetry (3) becomes a complex conjugation:

$$\phi \rightarrow \phi^* \quad \text{and} \quad \phi^* \rightarrow \phi , \quad (6)$$

and the discrete rotation (4) becomes a phase transformation:

$$\phi \rightarrow \phi \exp(i2\pi n/3) , \quad \phi^* \rightarrow \phi^* \exp(-i2\pi n/3) . \quad (7)$$

C. In polar coordinates the $Z(3)$ invariance is easily seen. The fields are $\phi_1 = \rho \cos \theta$, and $\phi_2 = \rho \sin \theta$. The potential part (2) of the Lagrangian becomes:

$$V(\rho, \theta) = \lambda \rho^4 - v \rho^3 \cos(3\theta) - \mu \rho^2 - \gamma . \quad (8)$$

The reflection transformation (3) becomes a reflection of the angle θ :

$$\theta \rightarrow -\theta , \quad (9)$$

and the rotational transformation (4) a translational operation:

$$\theta \rightarrow \theta + 2\pi n/3 , \quad n = 0, 1, 2 . \quad (10)$$

The only θ -dependent factor in (8) $\cos(3\theta)$ is evidently invariant under the mappings (9) and (10).

D. The last form is in terms of a matrix field that links the model to a $SU(2)$ representation. Define the matrix field:

$$\hat{\phi} = \begin{pmatrix} \phi_1 & -\phi_2 \\ -\phi_2 & -\phi_1 \end{pmatrix} = \sigma_3 \phi_1 - \sigma_1 \phi_2 \quad (11)$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices. Then we can write the Lagrangian density (1) as a trace (Tr) expression:

$$L = \frac{1}{4} \text{Tr} (\partial_\mu \hat{\phi})^2 - \frac{1}{2} \text{Tr} [\lambda \hat{\phi}^4 - v (\sigma_3 \hat{\phi})^3 - \mu \hat{\phi}^2 - \gamma I] , \quad (12)$$

I is the 2 x 2 dimensional unit matrix.

The reflection symmetry (3) becomes the mapping:

$$\hat{\phi} \rightarrow \sigma_3 \hat{\phi} \sigma_3, \quad (13)$$

and the Z(3) transformation the mapping:

$$\hat{\phi} \rightarrow O^+ \hat{\phi} O, \quad (14)$$

with the operator:

$$O = \exp(-i\pi n \sigma_2 / 3), \quad O^3 = I, \quad n = 0, 1, 2. \quad (15)$$

This matrix representation is a scalar version of the SU(2) Yang-Mills field⁶:

$$A_\mu = \sum_{a=1}^3 A_\mu^a \sigma_a,$$

with the imaginary component set to zero $A_\mu^2 = 0$. The representation (11) is suitable for generalization to more than two field components by an expansion in terms of the generators of the SU(3) group.

3. SOLITON SOLUTIONS

The Euler-Lagrange field equations in two dimensions are:

$$(a_0^2 - a_1^2)\phi_1 = -4\lambda(\phi_1^2 + \phi_2^2)\phi_1 + 3v(\phi_1^2 - \phi_2^2) + 2u\phi_1, \quad (16a)$$

$$(a_0^2 - a_1^2)\phi_2 = -4\lambda(\phi_1^2 + \phi_2^2)\phi_2 - 6v\phi_1\phi_2 + 2u\phi_2. \quad (16b)$$

First we are looking for time independent solutions of the field equations⁷. One dimensional solutions of the ϕ^4 theory have found applications in polyacetylene⁸. For further discussion of solitons in $(CH)_x$ see Ref. 9. In addition, references to this problem can be found in this volume of the proceedings.

Plane wave solutions in higher dimensional space can be obtained from the one dimensional solution by the mapping¹⁰ $x \rightarrow k_\mu x_\mu / k_\mu k_\mu$.

In order to solve the field equations, it is useful to consider the topology of the potential $V(\phi_1, \phi_2)$ as a surface defined on the ϕ_1, ϕ_2 space. There are three minima—also called vacua—denoted by $\Omega_1, \Omega_2, \Omega_3$, where the potential is zero, located on a circle of radius ϕ_V at an angle of 120° from each other. The coordinates of the minima are:

$$\Omega_1 = (\phi_V, 0); \quad \Omega_2 = (-\frac{1}{2}\phi_V, \frac{\sqrt{3}}{2}\phi_V); \quad \Omega_3 = (-\frac{1}{2}\phi_V, -\frac{\sqrt{3}}{2}\phi_V). \quad (17)$$

ϕ_V is the value of the vacuum field determined as the positive root of the equation $\partial V(\rho, 0) / \partial \rho = 0$. In the center of the ϕ_1, ϕ_2 plane the potential is elevated $V(0, 0) = -\gamma > 0$. For large values of the fields $\phi_1^2 + \phi_2^2 \gg \phi_V^2$ the potential has steep walls due to the positivity of λ . There are three saddle points located at $60^\circ, 180^\circ$ and 240° . The distance from the center of the saddle points is given by the positive root of the equation $\partial V(\rho, \pi/3) / \partial \rho = 0$.

Solutions tunneling from one vacuum to the other are obtained by the method of trajectories in field space¹¹.

When the parameters λ , v , μ , γ are related to each other in the following way:

$$\mu = \lambda \phi_V^2, \quad v = \frac{2}{3} \lambda \phi_V, \quad \gamma = -\frac{2}{3} \lambda \phi_V^4 = -v \phi_V^3, \quad (18)$$

the geometry of the potential arranges itself such that the saddle point between two vacua lies on the straight line connecting these two vacua. The tunneling from one vacuum to the other will now occur along this straight line through the saddle point.

There are six pairs of solutions to the field equations connecting the minima pairwise. Three pairs of solutions are generated by the mapping (4); the other three obtained by the reflection (3). We use the notation $\alpha = (3\lambda/2)^{1/2} \phi_V$. The basic set of solutions is:

1. The solution going from Ω_1 to Ω_2 , i.e. $\vec{\phi}(+\infty) \in \Omega_1$, $\vec{\phi}(-\infty) \in \Omega_2$ is

$$\phi_1(x) = \frac{1}{4} \phi_V [1 + 3 \tanh(\alpha x)]; \quad \phi_2(x) = \frac{\sqrt{3}}{4} \phi_V [1 - \tanh(\alpha x)] \quad (19)$$

The trajectory is the algebraic relation independent of x :

$$\phi_1 + \sqrt{3} \phi_2 = \phi_V \quad (20)$$

The next two pairs are obtained from (19) by rotation of 120° and 240° .

2. The solution tunneling from Ω_2 to Ω_3 with the asymptotic values $\vec{\phi}(+\infty) \in \Omega_2$, $\vec{\phi}(-\infty) \in \Omega_3$ is

$$\phi_1(x) = -\frac{1}{2} \phi_V; \quad \phi_2(x) = \frac{\sqrt{3}}{2} \phi_V \tanh(\alpha x) \quad (21)$$

The trajectory is

$$\phi_1 = -\frac{1}{2} \phi_V \quad (22)$$

3. The soliton tunneling from Ω_3 to Ω_1 $\vec{\phi}(+\infty) \in \Omega_3$, $\vec{\phi}(-\infty) \in \Omega_1$ is

$$\phi_1(x) = \frac{1}{4} \phi_V [1 - 3 \tanh(\alpha x)]; \quad \phi_2(x) = -\frac{\sqrt{3}}{4} \phi_V [1 + \tanh(\alpha x)] \quad (23)$$

with the trajectory

$$\phi_1 - \sqrt{3} \phi_2 = \phi_V \quad (24)$$

All trajectories together [Eqs. (20), (22), and (24)] form an equilateral triangle with the corners in the minima of the potential.

4. PROPERTIES OF THE SOLITON SOLUTIONS

The energy-momentum tensor is a functional of the fields and their derivatives defined as:

$$T_{\mu\nu} = \sum_{k=1}^2 \frac{\partial L}{\partial \phi_{k,\mu}} \phi_{k,\nu} - g_{\mu\nu} L; \quad \mu, \nu = 0, 1 \quad (25)$$

The kinetic and potential energy of the field configuration given by any of the Eqs. (19), (21), or (23) are equal. Their sum is the energy density:

$$T_{00}(x) = \frac{9}{8} \lambda \phi_V^4 [\cosh(\alpha x)]^{-4} . \quad (26)$$

The integrated field energy gives the mass of the soliton:

$$M = \int_{-\infty}^{\infty} dx T_{00}(x) = \alpha \phi_V^3 . \quad (27)$$

The other components of the field tensor—the field momentum T_{01} and the field pressure T_{11} —vanish.

The linear extension (analog of the radius in one dimension) of the energy distribution in configuration space is defined by:

$$R = [3 \int_0^{\infty} dx x T_{00}(x) / \int_0^{\infty} dx T_{00}(x)]^{1/2} , \quad (28)$$

$[T_{00}(x) = T_{00}(-x)]$, such that a box-like energy distribution of arbitrary height and length $2s$ gives the correct answer $R = s$. The dependence on λ and ϕ_V is explicit:

$$R = (0.80305\dots) \lambda^{-1/2} \phi_V^{-1} . \quad (29)$$

With $B = -\gamma > 0$ the mass of the soliton can be written in terms of the linear radius:

$$M = (1.1438\dots) B(2R) . \quad (30)$$

5. RESONANCES

Excited states of the scalar fields are given by fluctuations in time about the classical fields ϕ_1 and ϕ_2 . Again we consider the solution Eq. (19). The perturbative expansion of the fields¹² (Q^2 and Q^3 are neglected in the field equations):

$$\phi_1^*(t, x) = \phi_1(x) + Q(t, x) ; \quad \phi_2^*(t, x) = \phi_2(x) - \frac{1}{\sqrt{3}} Q(t, x) , \quad (31)$$

is chosen such that the new fields ϕ_1^* and ϕ_2^* satisfy the same algebraic relation (20) as the classical fields. With this expansion the energy spectrum of the resonance has two levels:

$$M^*(k) = M + E^{(k)} , \quad k = 1, 2 ; \quad (32)$$

where the energies and the perturbations are:

$$E^{(1)} = \sqrt{3} \alpha ; \quad Q^{(1)}(t, x) = A \cos(Et) \operatorname{sech}(\alpha x) \tanh(\alpha x) , \quad (33)$$

and

$$E^{(2)} = 2\alpha ; \quad Q^{(2)}(t, x) = A \cos(Et) \left[1 - \frac{3}{2} \operatorname{sech}^2(\alpha x) \right] . \quad (34)$$

6. TWO SOLITON INTERACTIONS

In physical applications the behavior of multicenter solutions is important. I have not found any exact multicenter solution in this model.

I will shortly analyze an ansatz for a field configuration with two centers at x_1 and x_2 . There is an ordering of the centers implied $x_2 \leq x_1$. The trial function contains linear^{13,14} and quadratic forms of the single soliton solutions Eq. (19) located at x_1 and the solution Eq. (23) with the sign of ϕ_2 changed [symmetry Eq. (3)] located at x_2 :

$$\begin{aligned} \phi_1^{(2)}(x, x_1, x_2) = & \cos^2(\epsilon) \frac{1}{4} \phi_V \left\{ 1 + 3[1 + \tanh(\alpha(x - x_1)) - \tanh(\alpha(x - x_2))] \right\} \\ & + \sin^2(\epsilon) \frac{1}{8} \phi_V [1 + 3\tanh(\alpha(x - x_1))][-1 + 3\tanh(\alpha(x - x_2))] , \end{aligned} \quad (35)$$

$$\begin{aligned} \phi_2^{(2)}(x, x_1, x_2) = & \cos^2(\epsilon) \frac{\sqrt{3}}{4} \phi_V \left\{ 1 - [1 + \tanh(\alpha(x - x_1)) - \tanh(\alpha(x - x_2))] \right\} \\ & + \sin^2(\epsilon) \frac{\sqrt{3}}{8} \phi_V [1 - \tanh(\alpha(x - x_1))][1 + \tanh(\alpha(x - x_2))] . \end{aligned} \quad (36)$$

ϵ is a variational parameter with values $\epsilon \in [0, \pi/2]$.

If one of the centers is removed to infinity, the fields (35,36) reduce to exact single soliton solutions. This limit process is independent of the value of ϵ . If $x_2 \rightarrow -\infty$, the fields (35,36) reduce to Eq. (19) and for $x_1 \rightarrow +\infty$ the fields reduce to Eq. (23) with the sign of ϕ_2 inverted.

The static potential energy between two solitons at a distance $d = x_1 - x_2$ is the difference between the total field energy and twice the mass of an isolated soliton:

$$P(d, \epsilon) = \int_{-\infty}^{\infty} dx T_{00}^{(2)}(x, d, \epsilon) - 2M . \quad (37)$$

A numerical calculation shows that the potential is negative, i.e. the two solitons attract each other when they come close together. For $\epsilon = 0$ (pure linear combination) we obtain a lower bound and for $\epsilon = \pi/2$ (pure quadratic form) an upper bound for the potential energy.

$$P(d, 0) \leq P(d, \epsilon) \leq P(d, \pi/2) \leq 0 ; \quad \epsilon \in (0, \pi/2) . \quad (38)$$

The potential converges fast to zero when the distance d becomes slightly larger than the linear extension of the particles $d > 2R$.

The quality of the ansatz Eqs. (35,36) can be checked by computing the action for this field configuration. The action per unit time is:

$$S(d, \epsilon) = - \int_{-\infty}^{\infty} dx T_{00}^{(2)}(x, d, \epsilon) . \quad (39)$$

For large distances $d \gg R$ the action is $-2M$ independent of ϵ . For fixed distances in the interaction region, the action has a minimum for $\epsilon = \pi/2$ and a maximum for $\epsilon = 0$.

$$-2M \leq S(d, \pi/2) \leq S(d, \epsilon) \leq S(d, 0) \leq 0 ; \quad \epsilon \in (0, \pi/2) . \quad (40)$$

In the static trial functions(35,36) the quadratic form of two single soliton solutions minimizes the action at any distance and seems therefore to be preferred to the linear form.

The net effect is that the static solitons attract each other. This is consistent with the findings in the ϕ^4 model at low velocities. High velocity solitons colliding with each other may scatter backward showing a repulsive potential at high speed^{14,15}.

CONCLUSIONS

We analyzed a field theory model with $Z(3)$ internal symmetry in $1 + 1$ dimensions. Exact soliton solutions are found in one space dimension. An ansatz for the two-center field shows that the static potential between two solitons is attractive.

ACKNOWLEDGMENTS

I am grateful to David K. Campbell for many helpful and interesting discussions.

This work was supported by the Director, Office of Energy Research, Division of Nuclear Physics of the Office of High Energy and Nuclear Physics of the U.S. Department of Energy under Contract W-7405-ENG-48.

REFERENCES

1. R.B. Potts, Some generalized order-disorder transformations, *Proc. Cambr. Phil. Soc.* 48 (1952) 106-109.
2. A.N. Berker, S. Ostlund and F.A. Putnam, Renormalization-group treatment of a Potts lattice gas for krypton adsorbed onto graphite, *Phys. Rev.* B17 (1978) 3650-3665.
3. A. de Rujula, H. Georgi and S.L. Glashow, A theory of flavor mixing, *Ann. of Phys.* 109 (1977) 258-313.
4. H.M. Ruck, Polynomial chromodynamics in 1+1 dimensions I. Confinement of quarks and the structure of composite particles, Lawrence Berkeley Laboratory preprint LBL-12316 (February 1981).
5. F. Constantinescu and H.M. Ruck, Quantum field theory Potts model, *J. Math. Phys.* 19 (1978) 2359-2361; Phase transitions in a continuous three states model with discrete gauge symmetry, *Ann. of Phys.* 115 (1978) 474-495.
6. C.N. Yang and R. Mills, Conservation of isotopic spin and isotopic gauge invariance, *Phys. Rev.* 96 (1954) 191-195.
7. H.M. Ruck, Solitons in cyclic symmetry field theories, *Nucl. Phys.* B167 (1980) 320-326.
8. M.J. Rice, Charged π -phase kinks in lightly doped polyacetylene, *Phys. Lett.* 71A (1979) 152-154.
9. W.P. Su, J.R. Schrieffer and A.J. Heeger, Solitons in polyacetylene, *Phys. Rev. Lett.* 42 (1979) 1698-1701; Soliton excitations in polyacetylene, *Phys. Rev.* B22 (1980) 2099-2111.
10. P.B. Burt, Solitary waves in nonlinear field theories, *Phys. Rev. Lett.* 32 (1974) 1080-1081.
11. R. Rajaraman, Solitons of coupled scalar field theories in two dimensions, *Phys. Rev. Lett.* 42 (1979) 200-204.
12. H.M. Ruck, Quark confinement in field theories with discrete gauge symmetry $Z(3)$, in: K.B. Wolf (ed.), *Group theoretical methods in physics*, Lecture Notes in Physics, 135 (Springer Verlag, Berlin, 1980).
13. A.E. Kudryavtsev, Solitonlike solutions for a Higgs scalar field, *JETP Lett.* 22 (1975) 82-83.
14. V.G. Makhankov, Dynamics of classical solitons (In non-integrable systems), *Phys. Rep.* C35 (1978) 1-128.
15. D.K. Campbell, private communication.

PART III
Reaction-Diffusion Processes

This Page Intentionally Left Blank

SOME CHARACTERISTIC NONLINEARITIES OF CHEMICAL REACTION ENGINEERING

Rutherford Aris

Sherman Fairchild Distinguished Scholar
California Institute of Technology

A brief survey is made of the different nonlinear forms which arise as expressions of chemical reaction rate in the general problem of diffusion and reaction in a solid catalyst. These may be classified in order of increasing complexity as powers and polynomials, rational functions, and transcendentials. The characteristics and some of the typical problems associated with each of these classes will be surveyed.

INTRODUCTION

In attempting to survey the characteristic types of nonlinearity of chemical reaction engineering it may be well to take examples from a general class of problems which is of central importance in the theory of chemical reaction processes.

The complexity of a catalytic system requires a very thorough analysis both of the physical and of the chemical processes that are at work. The porous catalyst pellet is a network of pore space of various and varying diameters often linked to a subsystem of even finer pores through a manufacturing process which involves the pressing together of a fine powder. No exact model of diffusion in so complicated a geometry is possible and, inevitably, a simplified model of the geometrical structure must be made. Even this is generally not taken to the second stage of being treated as a diffusion model with the reaction as in the boundary condition, but rather serves to define an effective diffusivity in which both the geometry of the pore structure and physical properties of the material which is diffusing are bound up together in an effective diffusivity. This diffusivity will generally be dependent on the concentration of the diffusing species but for many practical purposes this dependence is ignored. Jackson's monograph [10] gives an excellent treatment of this topic.

But, while it is often possible to make the assumption of constant diffusivity and conductivity, it is not usually feasible, or sensible, to make any simplifying assumption as to the reaction rate. This is generally present as a nonlinear function of the dependent variables in a pair of parabolic or elliptic equations. The full derivation of these will not be carried out in detail here for it is the burden of the second chapter of my treatise on the "Mathematical Theory of Diffusion and Reaction in Permeable Catalyst." ([4]; this will be referred to as MT, followed by a page or equation number.) Let it here suffice to say that the equations are obtained by the usual mass or energy balances and in their time-dependent form, give a pair of equations for the dimensionless concentration, u , and the dimensionless temperature, v . Let us further simplify the assumptions by assuming that only a single reaction is taking place and that this may be characterized by the concentration of a single component or by a variable which represents the extent of reaction.

The equation for the concentration then is

$$\frac{\partial u}{\partial t} = \nabla^2 u - \phi^2 R(u) \text{ in } \Omega, u=1 \text{ on } \partial\Omega \quad (1)$$

In this equation u is the concentration, made dimensionless by the surface concentration u_s , τ is the dimensionless time equal to a diffusion coefficient times the real time divided by the square of some length parameter; ϕ^2 is the ratio of the maximum rate of reaction to the corresponding diffusion flux and is a measure of the relative preponderance of reaction rate. $R(u)$ is the dimensionless reaction rate which at $u = 1$ has been normalized to have the value $R(1) = 1$. Again, for simplicity, this equation will be used with the Dirichlet condition that $u = 1$ on the surface of the region for which it is to be solved. In the equation itself the rate of change of concentration in any element is necessarily due to the net flux by diffusion (the first term on the right-hand side) minus the rate at which that substance is disappearing by the reaction. (Cf. MT. pp. 51-9.) When there are several reactions it is necessary to insert more variables and, of course, to have more than one reaction rate. When the reaction is not isothermal, i.e., the temperature of the catalyst pellet is not uniform or kept constant by some other means, it is necessary to make R a function of u and v .

A similar equation can be obtained by an energy balance in the particle and it yields a second partial differential equation for v , the dimensionless temperature. Again, this is a quasilinear parabolic equation in which the temperature change is the sum of a contribution from conduction diffusion and one from reaction:

$$L \frac{\partial v}{\partial \tau} = \nabla^2 v + \beta \phi^2 R(u, v) \text{ in } \Omega, v=1 \text{ on } \partial\Omega \quad (2)$$

Here β is a heat of reaction parameter and L the ratio of the thermal and material diffusivities. We will again keep things simple by using the Dirichlet conditions $v = 1$ on the boundary of the region. The partial differential equations obtain in a connected region of three dimensional space, Ω , with the boundary conditions asserted on the piece-wise smooth boundary of this region, $\partial\Omega$. (Cf. Fig. 1.)

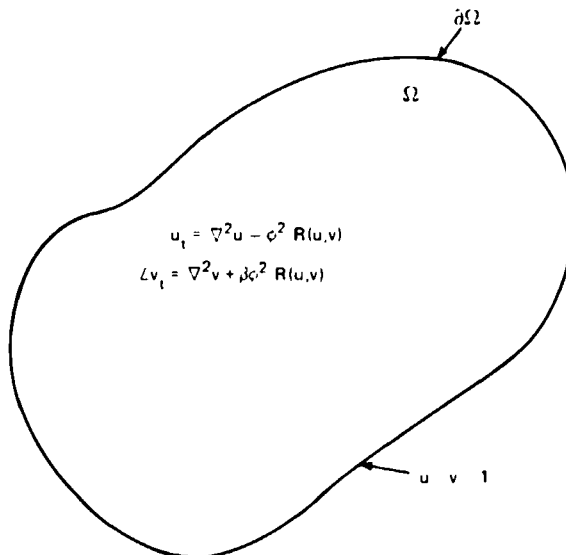


Figure 1: The catalyst pellet

The theme of the paper will, therefore, concern the behavior of this system when $R(u,v)$ takes on various forms. For the most part we shall be dealing with the steady state where the parabolic nature of the differential equations degenerates into ellipticity. Thus we will not be concerned with the initial conditions although there surely are such.

In the steady state there is a most important functional of the solution which represents the effectiveness of the catalyst pellet under the conditions of operation. It is, in fact, the average rate of reaction throughout the pellet divided by the rate of reaction which will obtain in the absence of any diffusion or conduction limitation whatsoever. Briefly, it can be given by the integral

$$\eta = \frac{1}{\bar{u}} \iiint_{\Omega} R(u,v) d\tau = \bar{R} \quad (3)$$

By its definition, this functional will tend to value 1 when the reaction is very slow compared with the speed of the diffusive processes and will tend to zero as the reaction rate dominates the possible rate of diffusion. The measure of this ratio of reaction to diffusion rate is the parameter ϕ , which is often referred to as the Thiele modulus in honor of the American engineer who, almost simultaneously with Zeldowich in Russia and Damköhler in Germany, recognized its importance and devised the effectiveness as a way of expressing its consequences. (For a fuller treatment of the history see MT pp. 37-42 and [3].)

Powers and Polynomials

Let us assume that the particle is isothermal and that we, therefore, have no need for the second equation derived from an energy balance. The simplest assumption is, of course, that the reaction is of the first order, which is expressed mathematically by making $R(u) = u$. This gives a thoroughly linear equation and it would be out of order in the present context to discuss it in any detail. One remark may be made in passing however, it concerns the simplest case and is almost trivially obvious. If the region Ω is a parallel-sided slab of infinite extent or sealed edges, then the equation becomes

$$\frac{d^2 u}{d\rho^2} = \phi^2 u \text{ in } (0,1), u(1) = 1, u'(0) = 0 \quad (4)$$

The immediate solution of this is

$$u(\rho) = \frac{\cosh \phi \rho}{\cosh \phi}, \quad \eta = \frac{\tanh \phi}{\phi} \quad (5)$$

A graph of η against ϕ in logarithmic coordinates shows the features that we should expect; namely, it is asymptotic to 1 as ϕ goes to 0 and to 0 as ϕ goes to infinity. This is a basic feature of the problem and these asymptotic relationships hold in a large number of other cases. In fact, the product of η and ϕ tends to a constant value and this is merely because at increasingly high reaction rates the diffusing and reacting substance disappears by reaction within a very short distance of the surface. The effectiveness of the pellet thus becomes proportional to its exterior surface area rather than to its volume. We remark in passing that there are some important generalizations of this linear case to more recondite shapes of Ω and to simultaneous first order reaction. The latter gives a rather similar formulation in terms of vectors of concentration and matrices representing diffusivities and reaction rate constants.

If the first-order reaction corresponds to the linear case and may be dismissed at a conference dedicated to nonlinear problems, the next easiest form of nonlinearity is surely that of the p th-order reaction which in symmetrical bodies (slab, cylinder or sphere) gives a form of the Emden-Fowler equation:

$$\nabla^2 u = \phi^2 u^p, \quad \frac{d}{dp} \left(\rho^q \frac{du}{dp} \right) = \phi^2 \rho^q u^p \quad (6)$$

In the geometry of the slab with $q = 0$ we see that the equation may be integrated by Clairaut's transformation i.e., by multiplying both sides by twice the first derivative and integrating. If M^2 is the first integral of the right-hand side of this equation,

$$[u'(p)]^2 = \phi^2 \int_{u_0}^u 2R(w)dw = \phi^2 [M(u; u_0)]^2 \quad (7)$$

Then, for the p th-order reaction we merely have some straightforward integrals to calculate.

$$M^2(u; u_0) = \frac{2}{p+1} u^{p+1} - u_0^{p+1} \quad (8)$$

$$\eta = \frac{1}{\phi} \left(\frac{2}{p+1} \right)^{\frac{1}{2}} \left\{ 1 - u_0^{p+1} \right\}^{\frac{1}{2}} \quad (9)$$

Whence a second integration gives

$$\phi p = \left(\frac{2}{p+1} \right)^{-\frac{1}{2}} \int_{u_0}^{u(p)} \left\{ u^{p+1} - u_0^{p+1} \right\}^{-\frac{1}{2}} du \quad (10)$$

and, since $u(1) = 1$,

$$\phi' = \left(\frac{2}{p+1} \right)^{-\frac{1}{2}} \frac{1}{u_0} \int_{u_0}^1 \left\{ u^{p+1} - u_0^{p+1} \right\}^{-\frac{1}{2}} du \quad (11)$$

It is best to turn this problem arsey-versy and let u_0 be the parameter in terms of which both the effectiveness factor and the Thiele modulus are calculated. Figure 2 shows the Thiele modulus as a function of the dimensionless center concentration and Figure 3 the effectiveness factor versus the normalized Thiele modulus. By normalized Thiele modulus we mean that the Thiele modulus is multiplied by a factor such that the asymptotic value of $\eta\phi$ is 1. It is clear from Eq. 9 that this is the square root of $(p+1)/2$. An interesting aspect of the second figure is that for values of p less than or equal to 1 the integral defining ϕ in terms of u_0 does not diverge as u_0 tends to zero. It follows that for values of ϕ greater than a certain critical value (itself dependent upon p) the only way in which a solution can be obtained is for there to be a dead region in the center of the pellet in which the concentration is identically zero. Then the boundary conditions

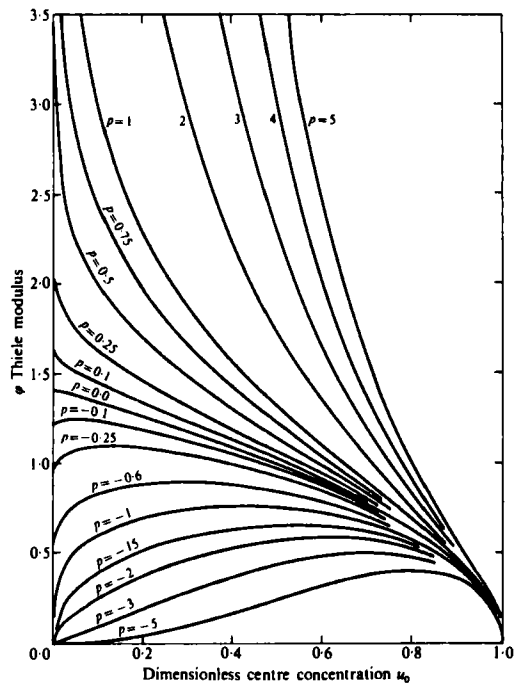


Figure 2: The relation between the dimensionless centre concentration, u_0 , and the Thiele modulus ϕ for the p th order reaction.

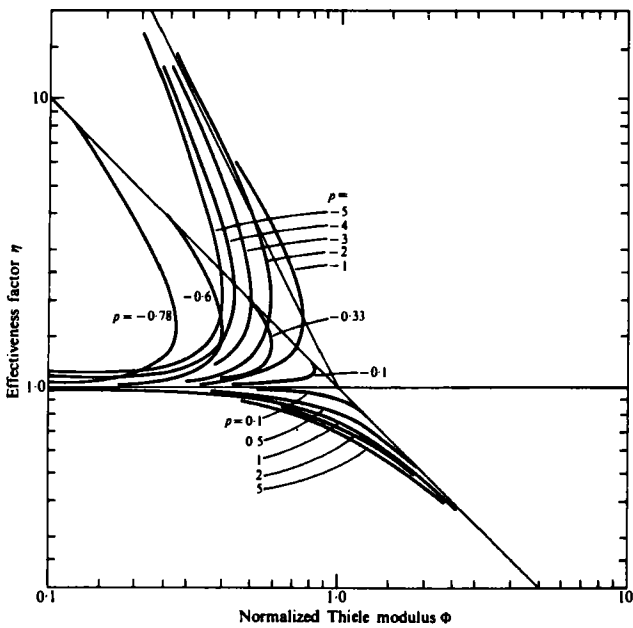


Figure 3: The effectiveness factor for the p th order reaction as a function of the normalized Thiele modulus.

$$\begin{aligned} u(\rho) &= u'(\rho) = 0 \quad \text{at } \rho = w \\ u(1) &= 1 \end{aligned} \quad (12)$$

serve to determine w as well as the constants of integration. For an interesting generalization of this free boundary phenomenon see Nichols' papers [13,14,15]; the geometry of a finite cylinder has been studied by Stephanopoulos [18].

The reduction to the hypergeometric function may be illustrated in cases $p > 1$ for then the substitution $U = 1 - (u_0/u)^{p+1}$ turns the integral in eq. (11) into an incomplete beta function and this can be expressed as

$$\phi = \left(\frac{2}{p+1}\right) u_0^{-(p-1)/2} (1-u_0^{p+1})^{1/2} F\left(\frac{1}{2}, \frac{1}{2} + \frac{1}{p+1}; \frac{3}{2}; 1-u_0^{p+1}\right) \quad (13)$$

Now the hypergeometric function converges on the circle $|1-u_0^{p+1}| = 1$, and it follows that since the exponent of u_0 is negative ϕ will range from infinity to zero as u_0 goes from 0 to 1. Also

$$\frac{d\phi}{du_0} = -\left(\frac{p+1}{2}\right)^{1/2} u_0^{-(p+1)/2} (1-u_0^{p+1})^{-1/2} F\left(-\frac{1}{2}, -\frac{1}{2} + \frac{1}{p+1}; \frac{1}{2}; 1-u_0^{p+1}\right) \quad (14)$$

showing that ϕ is monotonic decreasing and that the relation between ϕ and u_0 is unique. For $-1 < p < 1$ the hypergeometric function must be transformed by the relation

$$F(a, b; c; z) = (1-z)^{c-a-b} F(c-a, c-b; c; z) \quad ,$$

giving

$$\phi = \left(\frac{2}{p+1}\right)^{1/2} (1-u_0^{p+1})^{1/2} F\left(1, 1 - \frac{1}{p+1}; \frac{3}{2}; 1-u_0^{p+1}\right) \quad (15)$$

The solution for $p = -1$ may be obtained in terms of Dawson's integral

$$F(x) = e^{-x^2} \int_0^x e^{t^2} dt \quad , \quad (16)$$

and is

$$\phi = 2^{1/2} F\left[\left(\ln \frac{1}{u_0}\right)^{1/2}\right] \quad , \quad (17)$$

$$\eta = \left(2 \ln \frac{1}{u_0}\right)^{1/2} / \phi \quad (18)$$

Since F has a maximum of 0.541 the value of ϕ^* is 0.765.

The various cases can best be summarized in tabular form

$$m = 1/|p+1| \quad ,$$

$$\phi = (2m)^{\frac{1}{2}}(1-U) U^{\frac{1}{2}}F(\alpha, b; c; U) \quad , \quad (19)$$

$$\frac{d\phi}{du} = -(2m)^{-\frac{1}{2}}(1-U)^{\alpha'} U^{-\frac{1}{2}}F(\alpha', b'; c'; U) \quad .$$

When $p = -m/(m+1)$ or $-(2m+3)/(2m+1)$ and m is a positive integer the hypergeometric functions are polynomials of degree m .

Table 1

Range of p	m	U	α	a	b	c	α'	a'	b'	c'
$\infty, 1$	$0, \frac{1}{2}$	$1-u_0^{p+1}$	$m-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}+m$	$\frac{3}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$m-\frac{1}{2}$	$\frac{1}{2}$
$1, 0$	$\frac{1}{2}, 1$	$1-u_0^{p+1}$	0	1	$1-m$	$\frac{3}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$m-\frac{1}{2}$	$\frac{1}{2}$
$0, -\frac{1}{3}$	$1, \frac{3}{2}$	$1-u_0^{p+1}$	0	1	$1-m$	$\frac{3}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$m-\frac{1}{2}$	$\frac{1}{2}$
$-\frac{1}{3}, -1$	$\frac{3}{2}, \infty$	$1-u_0^{p+1}$	0	1	$1-m$	$\frac{3}{2}$	$1-m$	1	$1-m$	$\frac{1}{2}$
$-1, -\infty$	$\infty, 0$	$1-u_0^{-p-1}$	$\frac{1}{2}$	1	$\frac{1}{2}-m$	$\frac{1}{2}$	--	--	--	--

Rational Functions

The next most complicated nonlinearity arises in the case of kinetics which are governed by so-called Langmuir-Hinshelwood or Michaelis-Menten kinetics (actually the names of Hougen and Watson can be rightly associated with the engineering applications of this form of kinetics, as can those of Briggs and Haldane with the biochemical use). When the kinetics of the reaction depend on adsorption on the solid surface of the catalyst pellet it is natural and proper to make the rate of reaction a function of the adsorbed concentration rather than that of the gas and the pores. This can give a dimensionless rate of reaction of the form of a quotient of two polynomials:

$$R(u) = \frac{u(\varepsilon_0 + \varepsilon_1 u + \dots + u^{p-1})}{\varepsilon_0 + \varepsilon_1 + \dots + 1} \left(\frac{\kappa+1}{\kappa+u} \right)^q \quad (20)$$

A simple reduction allows this function of u to vanish at $u = 0$ as well as to satisfy the condition of being 1 at $u = 1$. In the slab geometry this yields a similar normalization of the Thiele modulus and leads to such curves as are seen in Figure 4 for various cases of the parameters. (For references see MT p. 170, Table 3.8.)

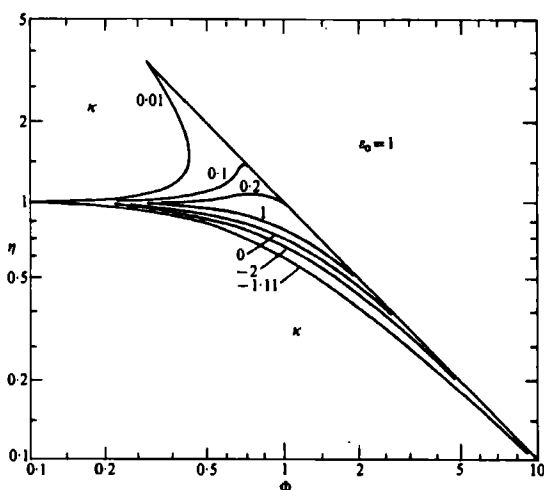


Figure 4: The normalized effectiveness factor plot for second-order Langmuir-Hinshelwood kinetics. Fixed $\Sigma_0=1$ and various κ .

The simplest form of the Langmuir isotherm, in which the adsorbed species tend to form a monolayer on the adsorbing surface, has some happily simple properties. In particular, in the simultaneous adsorption of two solids from a flowing medium the equations of balance are reducible and thus amenable to the hodograph transformation. Under their transformation the characteristics of the system in the plane of the two concentrations consists of a pattern of straight lines. This fortuitous linearity of a particular nonlinear system is the secret of the way in which it is possible to get some very far reaching results in the theory of chromatography of such substances. [17]

Transcendental Functions

Once the assumption of isothermality is abandoned it becomes inevitable that the kind of nonlinearity that will arise is a thoroughly transcendental one. This is because temperature influences the reaction rate constants in a highly nonlinear fashion so that they increase with temperature according to the law $d \ln k / dT = E/RT^2$. Thus, to take the very simplest of reactions, the irreversible p th-order disappearance of the key substance, the rate of reaction will be $k(T) c^p$ and $k(T) = k_0 e^{-E/RT}$. Reduction to dimensionless variables thus leads to a pair of equations of the form

$$\frac{\partial u}{\partial \tau} = \nabla^2 u - \phi^2 u^p \exp \gamma \left(\frac{v-1}{v} \right) \quad (21)$$

$$L \frac{\partial v}{\partial \tau} = \nabla^2 v + \beta \phi^2 \exp \gamma \left(\frac{v-1}{v} \right) \quad (22)$$

These equations are not as objectionable as one might at first suppose for the nonlinearity enters into one term only of the right-hand side and does so in precisely the same fashion for each. Thus if we form a combination of temperature and concentration

$$\beta u + v = z$$

and if $L = 1$, this combination satisfies the heat equation. In the steady state it satisfies Laplace's equation. But z is constant and equal to $(1+\beta)$ on the surface and hence, by the well-known property of a potential function, is constant everywhere.

Thus $\beta u + v \equiv 1 + \beta$ and

$$R(u, v) = G(u) = u^p \exp \left\{ \frac{\beta \gamma (1-u)}{1+\beta(1-u)} \right\} \quad (23)$$

or

$$R(u, v) = F(v) = \left(\frac{1+\beta-v}{\beta} \right)^p \exp \left\{ \gamma \frac{v-1}{v} \right\} \quad (24)$$

and

$$\nabla^2 u = \phi^2 G(u) \quad \text{or} \quad \nabla^2 v + \beta \phi^2 F(v) = 0 \quad (25)$$

If we choose u as the variable the various forms of $G(u)$ will, for constant p , depend on the two parameters β and γ . The same applies to the function $F(v)$ and may, perhaps, be illustrated rather more easily for this function. It has been normalized so that $F(1) = 1$ and since u vanishes when $v = 1 + \beta$ the curve must take one of the forms shown in Figure 5. In particular, if the point β, γ lies beneath the hyperbola $\beta\gamma = 1$ (i.e., in region A of the $\beta\gamma$ -plane) then the function F is monotone decreasing in the interval of interest which is $(1, 1+\beta)$. If (β, γ) lies in the region β then F has a maximum, but no point of inflection. On the other hand if β, γ lies in the region C between the two hyperbolae, $\beta\gamma/(1+\beta)$ is greater than 2 but less than 4, then the curve has an inflection point, but $F(v)/(v-1)$ is monotone decreasing as v passes from 1 to $1 + \beta$. However, if β, γ lies above this third hyperbola then F/v is not monotonic and Luss has shown that this is a necessary condition for multiplicity of steady states [12].

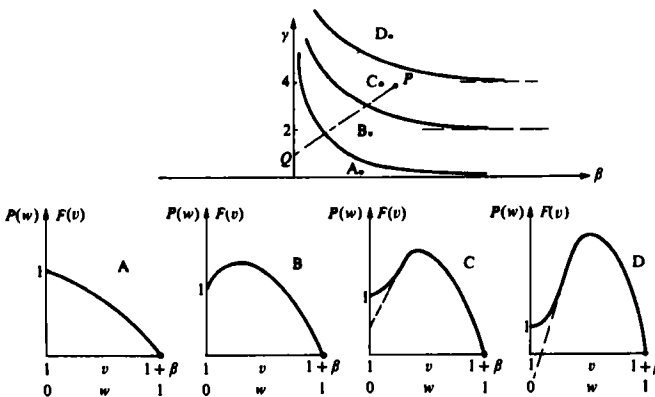


Figure 5: The forms of the function $F(v)$ and $P(w)$ for $p=1$ and several values of β, γ .

Let me turn aside to note that even here the nonlinear expression has a very interesting similarity property and one which allows every point of the calculation to give some information for a family of η, ϕ curves.

Villadsen and Michelsen [19] observed that the function $G(u)$ was patient of the following similarity transformation. If we write

$$G(u) = u^p \exp \left(\frac{\beta \gamma (1-u)}{1+\beta(1-u)} \right) = G(u; \beta, \gamma) \quad (26)$$

and set $u = u_1 U$, then after some algebraic manipulation we have

$$G(u; \beta, \gamma) = G(u_1; \beta, \gamma) G(U; B, \Gamma) \quad (27)$$

where

$$B = \beta u_1 / \{1 + \beta(1 - u_1)\} \quad , \quad (28)$$

$$\Gamma = \gamma / \{1 + \beta(1 - u_1)\} \quad .$$

Similarly, with the temperature function

$$F(v) = (1 + \beta - v)^p \exp(-\gamma/v) = F(v; \beta, \gamma) \quad (29)$$

we have for $v = v_1 V$

$$F(v; \beta, \gamma) = \frac{v_1^p}{1} F(V; B, \Gamma) \quad , \quad (30)$$

where

$$B = (1 + \beta - v_1) / v_1 \quad , \quad (31)$$

$$\Gamma = \gamma / v_1 \quad .$$

As we should expect, the loci of points B, Γ that are traced out in the (B, Γ) -plane by varying u_1 or v_1 are the same, namely

$$(1 + \beta)\Gamma - \gamma B = \gamma \quad . \quad (32)$$

This represents a straight line through (β, γ) and $(-1, 0)$.

This transformation implies that every point in the integration of the differential equations (25) for symmetrical geometries can be interpreted to give the effectiveness factor for some problem. Consider, for example, integrating the symmetrical equation (6). It is convenient to set $\xi = \phi p$ so that

$$\frac{d^2 u}{d\xi^2} + \frac{q}{\xi} \frac{du}{d\xi} = G(u) \quad (33)$$

and to integrate this as an initial value problem with

$$\frac{du}{d\xi} = 0, \quad u = u_0 \quad \text{at } \xi = 0 \quad . \quad (34)$$

When the point at which $u = 1$ is reached, the value of ξ is ϕ and

$$\eta = \frac{q+1}{\phi^2} \frac{du}{d\rho} = \frac{q+1}{\phi} \frac{du}{d\xi} = \frac{q+1}{\xi} \frac{du}{d\xi} \quad (35)$$

By keeping β and γ fixed and varying u_1 , the (η, ϕ) -curve can be traced out, for u_1 is a parameter along it. But if the integration is stopped at the point where $u = u_1$ a problem is solved for the parameters B and Γ given by eqn. (28). To see this we substitute

$$u = u_1 U, \quad \xi'^2 = \rho^2 \phi^2 G(u_1; \beta, \gamma) / u_1 \quad (36)$$

in eqn. (6) so that, by eqn. (27), it becomes

$$\frac{d^2 U}{d\xi'^2} + \frac{q}{\xi'} \frac{dU}{d\xi'} = G(U; B, \Gamma) \quad (37)$$

If this is integrated from $\xi' = 0$ until $U = 1$ then the value of ξ'^2 is

$$\phi^2 = \phi^2 \rho^2 G(u_1; \beta, \gamma) / u_1 \quad (38)$$

and

$$\eta = \frac{q+1}{\xi'} \frac{dU}{d\xi'} = \frac{q+1}{G(u_1; \beta, \gamma)} \frac{1}{\xi} \frac{du}{d\xi} \quad (39)$$

Thus the integration of eqn. (33) to the point at which $u = u_1$, say $u_1 = u(\xi_1)$, gives

$$\phi^2 = \xi_1^2 G(u_1; \beta, \gamma) / u_1 \quad (40)$$

$$\eta = \frac{q+1}{G(u_1; \beta, \gamma)} \left[\frac{1}{\xi_1} \left(\frac{du}{d\xi} \right) \right]$$

for the B and Γ given by eq. (28).

It has sometimes been felt that the nature of the nonlinearity in the reaction rate constant is mathematically complicated and that, since the usual range of temperatures is in the region where the exponential is increasing rapidly, it should be possible to replace the Arrhenius function by a positive exponential in temperature. When this is done, as is often the case in combustion studies, a very interesting equation results. In engineering circles this is associated with Franck-Kamenetskii [5] and in more mathematical with the name of Gelfand [7]. See also Joseph and Lundgren [11].

Though the justification for considering the problem

$$\nabla^2 u + e^u = 0 \quad \text{in } \Omega, \quad (41)$$

$$u = 0 \quad \text{on } \partial\Omega$$

may be rather slender if it is to be regarded as an approximation to the zeroth-order exothermic reaction, the equation remains of great interest. Frank-Kamenetskii (1939) was the first to apply it to combustion problems. He wrote the equation as

$$\rho^{-q} \frac{d}{dp} \rho^q \frac{du}{dp} + \delta e^u = 0 \quad , \quad (42)$$

$$u(\pm 1) = 0 \quad .$$

If the solution is computed in the usual fashion using u_0 as parameter there is a maximum value of δ , which Frank-Kamenetskii took to be the measure of a critical size above which spontaneous ignition should take place. This is true if spontaneous ignition is understood to mean that solutions of eqns. (42) only exist if there is a dead core with

$$u(\rho) = \beta\gamma \quad , \quad u'(\rho) = 0 \quad (0 \leq \rho \leq w) \quad . \quad (43)$$

This was first solved by quadratures for the slab ($q = 0$) and later Chambré (1952) found an analytical solution for the cylinder ($q = 1$). For the sphere no analytical solution is obtained, but the problem becomes particularly interesting as there are infinitely many solutions for a critical value of δ . The paper of Hlaváček and Marek [8] gives the most complete treatment from the chemical engineering viewpoint.

(i) The infinite flat plate. For the slab there is an analytical solution to eq. (42) namely

$$e^u = A \operatorname{sech}^2\{\rho\sqrt{(A\delta/2)}\} \quad , \quad (44)$$

and the boundary condition can be matched if

$$A = \cosh^2\{\sqrt{(A\delta/2)}\} \quad . \quad (45)$$

This provides an equation for A which has two solutions when

$$\delta < \delta_c = 0.878 \quad . \quad (46)$$

If $\delta = \delta_c$ there is just one solution with $u(0) = \ln A = 1.188$ and hence a maximum temperature rise within of $1.2RT^2/E$. This critical value of δ is held to be a criterion for the critical slab thickness above which no steady state solution can be found and is interpreted as the explosion limit.

However true this may be, in practice it is based on an incomplete analysis of the problem. Steady state solutions can be found for any value of δ but the boundary conditions (43) will have to be invoked for $\delta > \delta_c$. To see this we return to eqn. (42) and write the boundary conditions as

$$u(1) = 0 \quad , \quad u(w) = u_0 \quad , \quad u'(w) = 0 \quad . \quad (47)$$

Then if

$$w = 0 \quad , \quad 0 \geq u_0 \geq \beta\gamma \quad ,$$

we have eqn. (42) and if

$$1 > w > 0, \quad u_0 = \beta \gamma$$

we have eqn. (43). Multiplying eqn. (42) for $q = 0$ by $2u'(\rho)$ and integrating we have

$$[u'(\rho)]^2 = 2\delta(e^{u_0} - e^u) \quad (48)$$

and hence

$$\begin{aligned} (2\delta)^{\frac{1}{2}}(\rho - w) &= \int_{u(\rho)}^{u_0} \frac{du}{\{e^{u_0} - e^u\}^{\frac{1}{2}}} \\ &= 2e^{-\frac{1}{2}u_0} \operatorname{sech}^{-1}[\exp\{-\frac{1}{2}(u_0 - u(\rho))\}] \end{aligned} \quad (49)$$

The boundary condition at $\rho = 1$ is satisfied if

$$(2\delta)^{\frac{1}{2}}(1-w) = 2e^{-\frac{1}{2}u_0} \operatorname{sech}^{-1}(e^{-\frac{1}{2}u_0}) \quad (50)$$

which since $A = e^{u_0}$ is just a form of eq. (48) when $w = 0$. Let this equation be written

$$\delta(1-w)^2 = 2e^{-u_0} \{\operatorname{sech}^{-1}(e^{-\frac{1}{2}u_0})\}^2 \quad (51)$$

and observe the form of the curve relating $\delta(1-w)^2$ and u_0 as shown in Fig. 6.

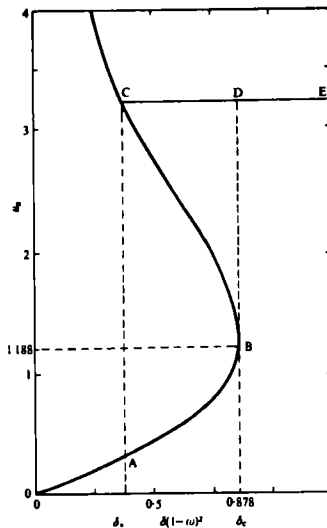


Figure 6: The solution of eq. (51).

Now the equation as we have written it in eq. (42) takes no account of the fact that a reaction cannot proceed when the reactant has been completely exhausted, i.e., when u exceeds $\beta\gamma$. Thus in place of e^u we should really have $e^{[1-H(u-\beta\gamma)]}$, where H is the Heaviside step function. If $\beta\gamma > 1.188$ the situation is as indicated by the upper horizontal line in Fig. 6 which is drawn through the point.

$$u_0 = \beta\gamma, \quad \delta(1-w)^2 = \delta_* = 2e^{-\beta\gamma} \{\operatorname{sech}^{-1}(e^{-\frac{1}{2}\beta\gamma})\}^2 \quad (52)$$

Since u_0 cannot exceed $\beta\gamma$ there is only one solution of eqn. (51) when $\delta < \delta_*$, namely a point on the part OA of the curve. There can be no solution with $w > 0$ for $\delta < \delta_*$. If $\delta_* < \delta < \delta_c$ there are two solutions with $w=0$ on the branches AB and BC respectively, but there is a third solution with $u_0 = \beta\gamma$ and $w > 0$ on the line CD. If $\delta > \delta_c$ then there is no steady state solution with $w=0$, but there is always a solution on DE with $w > 0$. In the latter cases the value of w is given by

$$w = 1 - (2e^{-\beta\gamma/\delta})^{\frac{1}{2}} \operatorname{sech}^{-1}(e^{-\frac{1}{2}\beta\gamma}) \quad (53)$$

If $\beta\gamma \leq 1.188$ then the horizontal line in Fig. 6 would have to be drawn from a point on the branch OB. In this case, the solution is unique. Hlaváček and Marek [8] appear to be the first to have recognized this feature of reactant exhaustion. The effectiveness factor

$$\eta = \int_w^1 e^u dp = -\frac{u'(1)}{\delta} = \left\{ \frac{2}{\delta} (e^{u_0} - 1) \right\}^{\frac{1}{2}} \quad (54)$$

when the solution satisfies the boundary conditions with $w > 0$ then

$$\eta = \left\{ \frac{2}{\delta} (e^{\beta\gamma} - 1) \right\}^{\frac{1}{2}} = \frac{1}{\phi} \left[\frac{2(e^{\beta\gamma} - 1)}{\beta\gamma} \right]^{\frac{1}{2}} \quad (55)$$

These curves are shown in Fig. 7, which is based on the calculations of Hlaváček and Marek. [9]

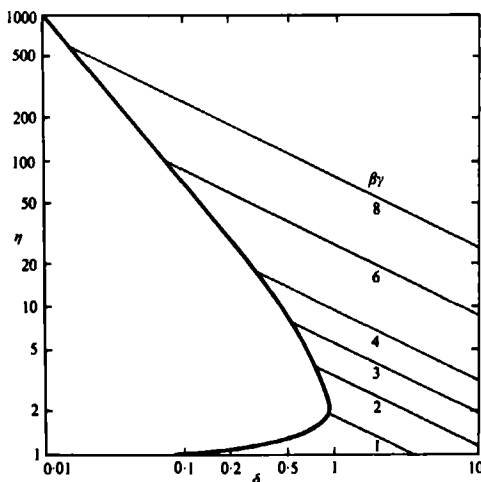


Figure 7: Effectiveness factor for Frank-Kamenetskii's equation in the slab.

(ii) The infinite cylinder. Chambré (1952) showed how to obtain an analytical solution of eq. (42) for the infinite cylinder, $q=1$. In case $w=0$ the substitutions

$$\rho \frac{du}{d\rho} = v, \quad \rho^2 e^u = w \quad (56)$$

are made, reducing the equation to

$$\frac{dv}{dw} + \frac{\delta}{2+v} = 0 \quad (57)$$

Since $v = w = 0$ at the center, the immediate integral of eq. (57) is

$$v^2 + 4v + 2\delta w = 0 \quad (58)$$

or

$$\rho^2 \left(\frac{du}{d\rho} \right)^2 + 4\rho \frac{du}{d\rho} + 2\delta \rho^2 e^u = 0 \quad (59)$$

But eq. (42) can be written

$$2\rho^2 \frac{d^2 u}{d\rho^2} + 2\rho \frac{du}{d\rho} + 2\delta \rho^2 e^u = 0 \quad (60)$$

so that by subtraction

$$\frac{d^2 u}{d\rho^2} - \frac{1}{\rho} \frac{du}{d\rho} - \left(\frac{du}{d\rho} \right) = 0 \quad (61)$$

The equation can be integrated to give

$$e^u = \frac{8B}{\delta(B\rho^2 + 1)^2} \quad (62)$$

where B has to be chosen to satisfy the boundary condition at the surface, i.e.

$$\delta(B+1)^2 = 8B \quad (63)$$

This equation again has two solutions provided that $\delta < \delta_c = 2$. The same type of analysis goes through the relation between u_0 and δ and is shown as one of the three curves in Fig. 8.

$$e^u = (\rho^*/b)^2 \cosh^2 \left[\ln \{ b + (b^2 - 1)^{1/2} \} + a \ln \rho \right] \quad (64)$$

$$w = e^{-1/2 \beta \gamma} \left\{ b^2 - \frac{2}{\delta} \right\}, \quad a = b(\delta/2)^{1/2} \quad (65)$$

where b satisfies

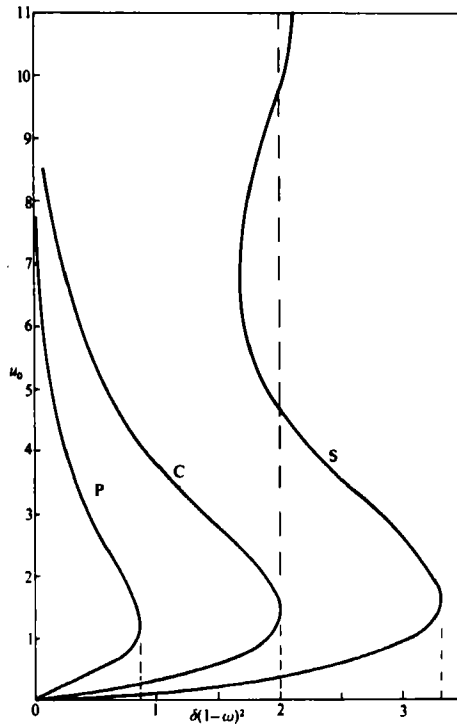


Figure 8: Centre 'temperature', u_0 , as a function of $\delta(1-\omega)^2$ for slab (L), cylinder (C), and sphere (S).

$$\{b+(b^2-1)^{1/2}\} \left[e^{-\frac{1}{2}\beta\gamma} \left\{ b^2 - \frac{2}{\delta} \right\}^{1/2} \right]^a = \frac{a-1}{a+1} \quad (66)$$

The effectiveness factor is then

$$\eta = \frac{4}{\delta} \left\{ 1 + \left(\frac{\delta}{2} \right) \sqrt{(b^2-1)} \right\} \quad (67)$$

(iii) The sphere. The integration of eq. (42) in the sphere requires, as we have seen in the previous section, a numerical integration and leads to the isothermal function or its generalizations. However, a very significant difference arises when the (u_0, δ) curve is continued to larger values of u_0 as in the curve S of Fig. 8. It is seen that the curve does not merely go through a maximum of δ , the so-called critical value $\delta_c = 3.32$, but that it does not come back asymptotically to the u_0 -axis but oscillates about the line $\delta=2$. Steggerda has given the first eight turning points:

δ	3.3220	1.6641	2.1083	1.9666	2.0095	1.9957	2.0002	1.9992
u_0	1.6075	6.7409	11.3769	16.1363	20.9159	25.3332	31.7950	35.6683

This means that we can obtain an arbitrarily large number of solutions by making $\beta\gamma$ large enough and δ close enough to 2. For $\delta=2$ there is a singular solution which we can obtain as follows. Let

$$v = \rho \frac{du}{d\rho} + 2, \quad w = u + 2 \ln \rho \quad (68)$$

then

$$\frac{dv}{dw} = \frac{2 - v - \delta e^2}{v} \quad (69)$$

This ordinary differential equation has a singular point at $v=0$, $w=\ln(2/\delta)$. But $w=\ln(2/\delta)$ is actually a solution of

$$e^u = \frac{2}{\delta \rho^2}, \quad (70)$$

which though it does not satisfy the boundary condition at $\rho=0$, will satisfy $u(1) = 0$ if $\delta = 2$. This is Emden's singular solution, and the solutions for $\delta = 2$ with u_0 corresponding to successive crossing points of the curve S of Fig. 8 with the vertical line $\delta = 2$ approaching more and more closely to the singular solution, departing only in the neighborhood of $\rho = 0$.

Since it is the sphere that first shows the phenomenon of infinite multiplicity, it is interesting to inquire what behavior the solution for a finite cylinder might exhibit. At one extreme - the coin - this is a flat plate; at the other - the curtain rod - it approaches the infinite cylinder. On the other hand, a cylinder with equal height and diameter has the same surface to volume ratio as a sphere and one is naturally led to ask if it gives rise to an infinity of solutions when the parameter is

$$\nabla^2 u + \delta e^u = 0$$

has some critical value.

The most complete results of this kind have been given by Villadsen and Ivanov [20] who treated the exponent q in eq. (42) as a "geometry factor." They showed that an infinite number of steady states prevailed when $1 < q < 9$. No more than three exist for $q \leq 1$ or $q > 10.44$ and only a finite number for $9 \leq q \leq 10.44$. Thus it would seem that when the pellet is the least bit more 'spherical' than it is 'cylindrical' (i.e., $q = 1+\epsilon$), there can be an infinite number of solutions.

An exact proof that the positive exponential approximation does not change the picture qualitatively has not been given for the case of diffusion and reaction though there are two directions from which it may be approached for the well-mixed reactor. Since there are grounds for believing that the qualitative behavior of the catalyst particle has some affinity with that of the stirred tank, it may be of interest to comment on this here [2]. One such discussion is to be found in Poore's paper in the Archive for Rational Mechanics and Analysis [16]. He studies the phase plane of the stirred tank and shows that it is closed in both for the Arrhenius function and the positive exponential.

Another approach has been provided by Golubitsky and Keyfitz [7], in their "qualitative study of the steady states solutions for a continuous flow stirred tank chemical reactor." Their argument turns on a careful analysis of the properties of the organizing center of this particular problem and they obtain formulae appropriate to a singularity they call the 'winged cusp.' Their arguments turn on the behavior of the Arrhenius expression and the inequalities which its derivatives satisfy. The positive exponential approximation could be obtained from their function by replacing γy by v and taking the limit of γ

going to infinity. The inequalities which they find necessary to impose are satisfied equally well by this limit as by the more complicated function. It would, therefore, seem clear that the qualitative behavior of the two systems is identical.

ACKNOWLEDGEMENT

This paper was prepared whilst the author had the privilege of being a Sherman Fairchild Scholar at the California Institute of Technology. Anyone who knows Caltech need only be told that in the generosity of its structure, this program fully reflects the excellence of that institution.

References

- [1] Amundson, N. R., and Schilson, R. E. (1961) "Intraparticle Diffusion and Conduction in Porous Catalysts," *Chem. Engrg. Sci.* 13, 226, 237.
- [2] Aris, R. (1969), "On Stability Criteria of Chemical Reaction Engineering," *Chem. Engrg. Sci.* 25, 149.
- [3] Aris, R. (1974), "The Theory of Diffusion and Reaction - A Chemical Engineering Symphony," *Chem. Engrg. Educ.* 8, 19.
- [4] Aris, R. (1975), "The Mathematical Theory of Diffusion and Reaction in Permeable Catalysts" (2 vols.), Clarendon Press, Oxford.
- [5] Frank-Kamenetskii, D. A. (1939), "Calculation of Thermal Explosion Limits," *Acta phys. - chim. URSS* 20, 365.
- [6] Gelfand, I. M. (1969), "Some Problems in the Theory of Quasilinear Equations," *Usp. Mat. Nauk* XV, No. 2, (86), 87. *Trans. Amer. Math. Soc.* (1963), (2), 29, 295.
- [7] Golubitsky, M., and Keyfitz, B. L. (1980), "A Qualitative Study of the Steady-State Solutions for a Continuous Stirred Tank Chemical Reactor," *SIAM J. Math. Anal.* 11, 316.
- [8] Hlaváček, V., and Marek, M. (1968), "Heat and Mass Transfer in a Porous Catalyst Particle. On the Multiplicity of Solutions for the Case of an Exothermic Zeroth-order reaction." *Coll. Czech. Chem. Commun. (Eng. Edn.)* 33, 506.
- [9] Hlaváček, V., and Marek, M. (1968), "Zum Entwurf katalytischer Rohrreaktoren bei radialen Wärmetransport," *Proc. Inst. Chem. Engrs. and V.T.G.-V.D.I. Meeting*, Brighton 1968.
- [10] Jackson, R., "Diffusion in Porous Media" (1977), Elsevier, New York.
- [11] Joseph, D. D., and Lundgren, T. S. (1973), "Quasilinear Dirichlet Problems Driven by Positive Sources," *Arch. Rational Mech. Anal.* 49, 241.
- [12] Luss, D. (1971), "Uniqueness criteria for lumped and distributed parameter chemically reacting systems," *Chem. Engrg. Sci.* 26, 1713.
- [13] Nicolaenko, B. (1977), "A General Class of Nonlinear Bifurcation Problems from a Point in the Essential Spectrum. Application to Shock Wave Solutions of Kinetic Equations," *Applications of Bifurcation Theory*, Academic Press, New York, 1977.

- [14] Nicolaenko, B., and Brauner, C. M. (1979), "Singular Perturbations and Free Boundary Value Problems," Proc. IVth International Symposium on Computing Methods in Applied Sciences and Engineering. I.R.I.A. Versailles, 1979, North-Holland Publishing.
- [15] Nicolaenko, B. and Brauner, C. M. (1979) "Nonlinear Eigenvalue Problems which Extend into Free Boundary Problems," Proc. Conf. on Nonlinear Eigenvalues. Springer-Verlag Lecture Notes in Mathematics No. 782.
- [16] Poore, A. B. (1973), "Multiplicity, Stability and Bifurcation of Periodic Solutions in Problems arising from Chemical Reactor Theory," Arch. Rational Mech. Anal. 52, 358.
- [17] Rhee, H-K., Aris, R., and Amundson, N. R. (1971), "Multicomponent Adsorption in Continuous Countercurrent Exchangers," Phil. Trans. Roy. Soc. A. 269, 187.
- [18] Stephanopoulos, G. (1980), "Zero Concentration Surfaces in a Cylindrical Catalyst Pellet," Chem. Engrg. Sci. 35, 2345.
- [19] Villadsen, J., and Michelsen, M. L. (1972), "Diffusion and Reaction on Spherical Catalyst Pellets: Steady-state and Local Stability Analysis," Chem. Engrg. Sci. 27, 751.
- [20] Villadsen, J., and Ivanov, E. (1978), "A note on the multiplicity of the Weisz-Hicks problem for an arbitrary geometry factor," Chem. Engrg. Sci. 33, 41.

This Page Intentionally Left Blank

PROPAGATING FRONTS IN REACTIVE MEDIA

Paul C. Fife
Mathematics Department
University of Arizona
Tucson, AZ 85721

Wavelike phenomena of many different types can be modeled with the aid of systems of partial differential equations of convection-diffusion-reaction type. Combustion theory provides a typical setting for these equations, but they are used in a great number of other contexts as well. I shall concentrate on one-dimensional steady-state traveling front solutions. A review (necessarily incomplete) will be made of a number of important advances in (1) asymptotic and other methods for reducing the complexity of given systems, and (2) more rigorous results on systems of one or two equations.

INTRODUCTION

The subject of my talk, propagating fronts, has to do with a phenomenon observed in a wide variety of contexts - physical, chemical, and biological. The analogy is more than superficial: to a great extent, similar concepts and tools can be used both in modeling and in analyzing wave phenomena in all these disciplines. I shall of course touch on some, possibly the most basic, of the "common ground" in this lecture; but much must necessarily be left unsaid.

First, the general picture. The prototypical example of fronts in reactive media is a flame, and a couple of my case studies will be drawn from flame theory. But more abstractly, one can simply envisage a medium continuously distributed in space which can exist, at each point in space, in a variety of possible states, described by an n -dimensional vector. These states are dynamic, in that they can change according to known rules. We assume there are at least two rest states available; these are states which will not change unless perturbed by some influence beyond the local dynamics.

Suppose the entire medium exists initially in one of the available rest states, and that a local "ignition" event happens: in a limited region of space, the medium is taken into the domain of attraction of another rest state. This perturbation, and the resultant attraction to the other state, may affect nearby regions in the medium through some transport mechanism, and these nearby regions are induced to undergo a similar transition between rest states. This is the beginning of a chain reaction, or domino effect. Let us call it a propagating front. In the above picture, a front will propagate in all directions from the ignition spot, but I would like to limit the concept to the case of a unidirectional front in a one-dimensional medium. With x denoting position and $u(x,t) \in \mathbb{R}^n$ the state of the system, the state dynamics and transport mechanism are assumed to be described by a system of partial differential equations of the type

$$u_t + (f(u))_x = (D(u)u_x)_x + g(u) \quad (1)$$

(D is a matrix). Deterministic continuous space-time models of the situation described above almost always may be put into either this form or that of an integro-differential equation. The second term on the left of (1) accounts for "convection", and those on the right are the "diffusive transport" and "reaction" terms. We call D the diffusion matrix.

Fronts are now defined to be solutions of the form $u = u(x-ct) = u(z)$ for some velocity c , satisfying

$$u(\pm\infty) = u_{\pm} \quad (\text{bounded distinct limits})$$

The front phenomenon is essentially nonlinear, because no such solutions exist when f and g are linear functions of u (at least when D is a constant).

I shall consider only smooth solutions (and so will exclude strictly discontinuous shocks). Front solutions satisfy

$$-cu_z + (f(u))_z = (D(u)u_z)_z + g(u) \quad (2a)$$

$$u(\pm\infty) = u_{\pm} \quad (2b)$$

The following basic questions about fronts arise:

- a. For which u_+ , u_- , and c do there exist fronts?
- b. What is the structure of the front's profile?
- c. Is the front stable,
 - (i) in one space dimension
 - (ii) when imbedded in a higher dimensional space?
- d. If unstable, what other stable structures may appear?

The emphasis here will be on (a) and (b), and to some extent (c). The talk will be in two parts:

- I. Techniques (primarily asymptotic) to reduce the complexity of systems.
- II. Some of the more detailed and rigorous results that have been obtained for the simplest systems (one or two equations).

I have not attempted to compile complete reference lists for results similar to those described here, or even to trace the origin of many of the ideas outlined. Moreover, the vast amount of effort on numerical approaches will not be mentioned further.

There are certain immediate necessary conditions for the existence of a front which should be made clear in these introductory comments. Integrating (2a) from $-\infty$ to ∞ , one obtains

$$-c\Delta u + \Delta f(u) = \int_{-\infty}^{\infty} g(u(z))dz \quad (3)$$

where $\Delta u = u_+ - u_-$. This is a condition relating u_+ , u_- , and c , but it is not too useful if the integrand $g(u(z))$ is not known. A typical occurrence is that some components of g vanish. I shall, in fact, segregate the components of u accordingly:

u_1 is a physical variable if $g_1(u) \equiv 0$;

u_1 is a chemical variable if $g_1(u) \equiv 0$.

(The number of each kind is not an invariant with respect to transformations of (1) into equivalent systems.) The set of physical variables I denote by \hat{u} , the chemical ones by \tilde{u} ; then

$$u = (\hat{u}, \tilde{u}), \text{ and similarly } f = (\hat{f}, \tilde{f}).$$

From (3) we obtain

$$c\Delta u = \Delta f(u) \quad (\text{Rankine-Hugoniot relations}) \quad (4)$$

and

$$g(u_{\pm}) = 0 \quad (5)$$

In all, these are n equations relating u_{\pm} and c . Though necessary, their satisfaction is by no means sufficient for the existence of a front.

1. REDUCING A SYSTEM'S COMPLEXITY

I shall proceed by outlining four examples of asymptotic methods by which larger systems may be reduced to a combination of smaller systems. This not only results in an easier analysis of the original system, but it also clarifies some qualitative features of the front's profile. These four examples are meant to exemplify four categories of approaches, each really comprising many possible variations or relatives which will not all be mentioned. And indeed in some cases many significant related treatments have appeared in the literature. Again, my intention is not to give a complete survey.

I shall attempt to disregard details which do not bear directly on the mathematical structure of the argument being presented; this is done in the interest of saving time and of elucidating that basic structure itself by doing away with confusing side issues. For example, I shall not point out the various dimensionless constants usually associated with combustion problems.

Finally, the situations described involve a small parameter ε , and the methods can be classed as lowest order formal asymptotic methods as $\varepsilon \rightarrow 0$. Approximations which are higher order in ε can be obtained, though sometimes with difficulty.

1. ZND and CJ detonation analysis.

This is the oldest and simplest of the examples. The names associated with the initials are Zeldovič, von Neumann, Doering, Chapman, and Jouguet. The first three developed their approaches to detonations independently of one another, around the time of World War II. The CJ models go back further. There are several possible avenues to the study of ZND detonations (see Williams (1965), for example). The approach I take is simply to assume the transport matrix D is small, and proceed by formal asymptotics.

Therefore replace D by εD in (2), where $\varepsilon \ll 1$. A front can be constructed in three parts: an abrupt part (shock) at $z = 0$ separating a fore and an aft region. We assume u_+ (the state of the medium into which the front advances) to be given, and the determination of c and u_- to be part of the problem.

(1) Ahead of the shock, set

$$u \equiv u_+, \quad z > 0.$$

Clearly, by (5) this constant is a solution of (2a).

(2) The shock has profile determined by stretching the variable z . Set $\zeta = z/\varepsilon$, $u(z) = U(\zeta)$, so that

$$-cU_{\zeta} + (f(U))_{\zeta} = (D(U)U_{\zeta})_{\zeta} + O(\varepsilon), \quad (6a)$$

$$U(\infty) = u_+. \quad (6b)$$

Neglecting the $O(\epsilon)$ correction, one discovers that the reaction terms have vanished. The reduced problem is therefore like a physical variable problem, and the Rankine-Hugoniot necessary conditions apply:

$$c(U_+ - U_-) + f(U_+) - f(U_-) = 0. \quad (7)$$

Since $U_+ = u_+$ is known, this is an equation relating U_- to c . Let us assume that there exists at least one solution U_- of (7) for each c in some parameter range I , and moreover that there exists an exact solution of (6) connecting U_+ to U_- . Then there will be a one-parameter family of possible shock profiles $U(\zeta)$, with velocity c being the parameter.

In cases of most physical interest, the system (7) can be reduced to the well-known Rankine-Hugoniot system in the physical variables density, momentum, and energy. In fact, assume that these are the physical variables in question, and that the chemical variables are given by $\hat{U} = \rho Z$ with $\hat{f} = \rho v Z$, where Z is the vector of mass fractions of the chemical species, v is the fluid velocity; and ρ is the total density. If $\rho = \hat{U}_1$, say, then $\hat{f}_1 = \rho v$. Let us write $f(U)$ as a function of \hat{U} and Z : $f(\hat{U}, Z)$. The following fact can be easily verified: If, for some constant Z_+ , the vectors \hat{U}_\pm satisfy

$$c\Delta\hat{U} + \Delta f(\hat{U}, Z_+) = 0, \quad (8)$$

then (7) is satisfied with these given \hat{U}_\pm , and with $\tilde{U}_\pm = \rho_\pm Z_+$.

On the Rankine-Hugoniot level, then, the problem reduces to (8) with Z_+ given by the conditions ahead of the shock. This is an elementary and well-studied problem. On the higher level of the full equations (6), let us simply assume the existence of a solution connecting U_+ and U_- . If there is no species diffusion (more specifically, if the chemical components of the vector DU_ζ vanish whenever $Z = \text{constant}$), then such an existence result follows from a result of Gilbarg (1951), to be mentioned later. All this results in a one-parameter family of shock profiles $U(\zeta)$ with $U(-\infty) = U_-$, depending on the parameter c .

(3) Behind the shock, we revert to the original variables, so (6a) applies with D replaced by ϵD . This equation is to hold for $z < 0$. To match with the shock profile considered above, it is necessary that $u(0) = U_-$. Now set $\epsilon = 0$ to obtain an ordinary differential equation initial value problem:

$$-cu_z + (f(u))_z = g(u), \quad z < 0 \quad (9a)$$

$$u(0) = U_- . \quad (9b)$$

Since g figures into this problem, the region behind the shock is the reaction zone. In typical cases, the solution of this problem will approach a rest state u_∞ as $z \rightarrow -\infty$. This final state depends on c , but in any case will satisfy (4) and (5).

In summary, the problem of constructing a front has been reduced to

- (1) a shock problem for the physical variables only, plus
- (2) an initial value problem for a first order ordinary differential equation in a reaction zone behind the shock.

This generally yields a one-parameter family of fronts. The next problem would be to decide which one of these actually occurs; the answer to this question depends on other aspects of the phenomenon (experimental conditions, etc.) being modelled. These other aspects, in fact, usually do serve to select a unique relevant candidate from among the collection we have constructed. Loosely

speaking, if there are no extra physical constraints imposed at $z = -\infty$, then the so-called Chapman-Jouguet values of u_- and c are the relevant ones.

A further simplification of this analysis produces the easiest of the detonation models: that of Chapman and Jouguet. For this one assumes the reaction rate g is large, but not so large as to overshadow the smallness of the transport term. Then replace g by kg , where $k \gg 1$ but $kc \ll 1$. The only effect of this in the above analysis is to accelerate the equilibration process behind the shock. In fact, with this parameter k introduced, one may define a new space coordinate $\eta = kz$, so that (9a) becomes

$$-cu_\eta + (f(u))_\eta = g(u), \quad \eta < 0.$$

Equilibration will occur over a distance in the variable η , of the order of magnitude 1, hence a distance, in z , of the order $1/k \ll 1$. The Chapman-Jouguet model takes this distance to be 0, so that the reaction zone and shock are both compressed to a single point.

2. High activation energy asymptotics for flames and chemical reactors.

Asymptotics of this sort were apparently first used (independently) by Bush-Fendell (1970) and Lifan (1971). Many people, including Buckmaster, Kapila, Ludford, Margolis, Matkowski, Sivashinsky, van Harten and Williams, have since done notable work in the area.

The most essential assumptions on which the method is based involve the strong dependence of the reaction function g on temperature, which is taken to be one of the "chemical" variables. This dependence is expressed by writing

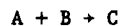
$$g = g\left(\frac{u}{\epsilon}, \epsilon\right), \quad \epsilon \ll 1. \quad (10)$$

This, of course, says nothing in itself, because any function can be written this way. Our main assumption concerns the character of the dependence of g on the two arguments, and will be spelled out later. The form (10) suggests that g , in fact, depends strongly on all the other state variables as well. Again, this form involves no loss of generality because of the (as yet) arbitrary dependence of g on the second variable, and I use it for convenience.

My description of the method proceeds under several inessential (or only partly essential) assumptions, besides the principal one. The method can be extended to cases when these inessential ones, which are listed below, are relaxed to some extent.

(1) The equations for the physical variables \hat{u} do not depend on \tilde{u} . This assumption means, for instance, that the heat generated by the chemical reaction does not affect the density very much. Under these conditions, $\hat{u} = \text{constant}$ will be a solution of the physical equations, and that is the solution we choose. Since \hat{u} is a constant, we disregard any dependence on \hat{u} in the considerations below. Thus $u = \tilde{u}$ consists only of chemical variables.

(2) The chemistry results from a single exothermic reaction



Besides the temperature T , the only relevant chemical variables are the concentrations c_A and c_B of A and B . The state vector u has three components: $(u_1, u_2, u_3) = (T, c_A, c_B)$. Furthermore, the single reaction results in a "g" of the special form

$$g\left(\frac{u}{\epsilon}, \epsilon\right) = \phi\left(\frac{u}{\epsilon}, \epsilon\right)K,$$

where ϕ is a scalar rate function and K a constant 3-vector.

(3) D is invertible and positive definite. (It is not necessary that D be diagonal, or even constant.)

(4) $f = \tilde{f} = 0$. Usually $f_1 = qu_1$, q being the velocity, supposed constant in this case. The coefficient q_1 can be absorbed into the definition of c , so the fourth assumption involves no great loss of generality.

(5) ϕ vanishes if and only if c_A or c_B is zero. (Mathematically, it may be convenient to allow negative concentrations; then we assume $\phi = 0$ for $c_A < 0$ or $c_B < 0$).

It will turn out that u_- is completely determined from the knowledge of u_+ and the assumption that at $z = -\infty$, the reaction has gone to completion in the sense that one of the c_A or c_B is zero, the other > 0 . I shall return to this part later; for the moment, assume the u_+ is prescribed and u_- thereby determined (we have yet to find c).

The equations are now

$$-cu_z = (Du_z)_z + \phi\left(\frac{u}{\epsilon}, \epsilon\right)K \quad (11)$$

Preliminary to the analysis, one rescales so that the reaction function is the correct order of magnitude. For this, it is appropriate (i) to express ϕ in terms of $u - u_-$ rather than u , and (ii) to formally separate out the ϵ -dependence as follows:

$$\phi\left(\frac{u}{\epsilon}, \epsilon\right) = \eta(\epsilon) \frac{1}{\epsilon} \psi\left(\frac{u-u_-}{\epsilon}\right)(1 + o(1)) \quad (\epsilon \rightarrow 0). \quad (12)$$

In typical cases, this separation is straightforward and defines ψ and η uniquely, up to order of magnitude. The details for Arrhenius temperature dependence will be given at the end of this section. Now define

$$z' = \eta^{1/2} z, \quad c' = \eta^{-1/2} c, \quad \text{drop the primes, and neglect the } o(1) \text{ term in (12):}$$

(11) becomes

$$-cu_z = (Du_z)_z + \frac{1}{\epsilon} \psi\left(\frac{u-u_-}{\epsilon}\right)K. \quad (13)$$

This is the basic equation to be analyzed.

We now come to the heart of high activation energy asymptotics, and the main assumption:

Essential assumption: For any state u_0 , componentwise strictly between u_- and u_+ ,

$$(i) \quad \lim_{s \rightarrow \infty} s\psi((u_0 - u_-)s) = 0 \quad (14)$$

and

$$(ii) \quad 0 < \int_0^\infty \psi((u_0 - u_-)s)ds < \infty. \quad (15)$$

An intuitive explanation is needed. At $z = -\infty$, the gas is burned, so c_A or $c_B = 0$; i.e., u_{2-} or $u_{3-} = 0$. Hence according to assumption (5) above, $\psi((u_0 - u_-)s) = 0$ for $s < 0$. For $s > 0$, $\psi((u_0 - u_-)s)$ is partly a measure of the rate of the reaction in its dependence on temperature, relative to its rate at the maximum temperature u_{1-} . If one thinks of $s = 1/\epsilon$, (14) and (15) simply say that for fixed temperature less than the maximum, this relative rate falls off "rapidly enough" as $\epsilon \rightarrow 0$.

The front is constructed in three parts.

(a) Since ϕ is zero when c_A or $c_B = 0$, and this is the case when $u = u_-$, it is true that $\phi(\frac{u_-}{\epsilon}, \epsilon) = 0$. Therefore the constant $u = u_-$ (burned gas) is a solution of (13). Take this as our solution for $z < 0$.

(b) Near $z = 0$, there is burning. It occurs mainly near $u = u_-$, because of the high activation energy. Going under the assumption that the flame is narrow, one stretches variables to analyze its internal structure:

$$\zeta = \frac{z}{\epsilon}, \quad v = \frac{u - u_-}{\epsilon}.$$

In these new variables, the equation (13) becomes, to lowest order,

$$D_0 v_{\zeta\zeta} + \psi(v)K = 0, \quad v(-\infty) = 0, \quad (16)$$

where $D_0 = D(u_-)$. There are two linearly independent vectors M_1, M_2 perpendicular to K . Let $w_i = M_i \cdot Dv$. Then (16) implies

$$w_{i\zeta\zeta} = 0,$$

and since $w_i(-\infty) = 0$, it follows that $w_i \equiv 0, i=1,2$. Inside the flame, therefore, there are two linearly independent relations satisfied by the components of v . This means that the vector v is collinear with a single constant vector V , which is perpendicular to both the vectors $M_1 D_0$. Clearly V may be taken to be the vector $V = D_0^{-1} K$. Thus

$$v = a(\zeta)V. \quad (17)$$

When (17) is substituted into (16), a single equation for a emerges:

$$a'' + \psi(aV) = 0, \quad a(-\infty) = 0 \quad (18)$$

This determines $a(\zeta)$ uniquely, up to translation, for $-\infty < \zeta < \infty$. Of greatest interest is $a'(\infty)$, which can be found by multiplying (18) by a' and integrating from $-\infty$ to ∞ :

$$a'(\infty) = \left(2 \int_{-\infty}^{\infty} \psi(sV) ds \right)^{1/2} \equiv \gamma. \quad (19)$$

This, together with (17), gives the jump discontinuity suffered by the derivative of u across the narrow flame. Specifically,

$$u_z(0+) = v_{\zeta}(\infty) = a'(\infty)V = \gamma V. \quad (20)$$

(c) For $z > 0$, we revert to (13). Since in this region u is strictly between u_- and u_+ , assumption (14) says that the reaction term in (13) is negligible for small ϵ . There results the following boundary value problem:

$$(Du_z)_z + cu_z = 0, \quad z > 0; \quad (21a)$$

$$u(0) = u_-, \quad u(\infty) = u_+, \quad (21b)$$

$$u_z(0) = \gamma V, \quad (21c)$$

the last coming from (20). There being too many boundary conditions, this problem is overdetermined and has a solution only if a compatibility condition is satisfied. To determine what that condition is, integrate (21a) once to obtain

$$Du_z + cu = \text{const} = cu_+.$$

At $z = 0$, this gives

$$D_0 u_z(0) = c(u_+ - u_-).$$

In view of (21c), the condition is therefore

$$\gamma D_0 V = c(u_+ - u_-). \quad (22)$$

Although this is a vector equation, it turns out that there is a single value of c for which it is satisfied. To see this, take the scalar product of (13) with the two vectors M_1 orthogonal to K . After integrating once, one finds

$$cM_1 \cdot u + M_1 \cdot Du_z = \text{const.}$$

The values at $z = \pm \infty$ must be the same, and $u_z = 0$ there, so $M_1 \cdot (u_+ - u_-) = 0$, which means that $(u_+ - u_-)$ is a multiple of K :

$$(u_+ - u_-) = bK, \quad (23)$$

Recall that by definition, $D_0 V = K$. Therefore (22) becomes $\gamma = cb$. This determines the velocity c .

With c known, the profile u for $z > 0$ can be determined by solving

$$u_z + cD^{-1}(u)(u - u_+) = 0, \quad u(0) = u_-.$$

The positive definiteness of D and positivity of c ensure the existence of a solution with $u(\infty) = u_+$.

At this point I would like to do two things: go back and justify the claim that u_- can be determined beforehand, and secondly go through the above construction of ψ for Arrhenius temperature dependence.

The first is easy. From (23) it is clear that to determine u_- , all we need is b . But this can be determined from the condition that one of the two concentrations $c_A = u_2$ or $c_B = u_3$ is zero. Suppose that $u_{2+} < u_{3+}$. Then it is u_2 that vanishes at $-\infty$, so the second component of (23) says $b = -u_{2-}/K_2$, and we are done.

For Arrhenius kinetics, the reaction function can be taken as

$$\phi = \alpha(\epsilon) c_A c_B \exp[-1/\epsilon \hat{T}], \quad (24)$$

where \hat{T} is absolute temperature, and α may depend on ϵ to account for its possibly being large or small. Let \hat{T}_+ be the temperature of the incoming gas, and define a scaled temperature

$$T = \frac{\hat{T} - \hat{T}_+}{\hat{T}_+} \equiv u_1.$$

The explicit dependence of ϕ on $\frac{u_1}{\epsilon}$, $\frac{u_2}{\epsilon} = \frac{c_A}{\epsilon}$, and $\frac{u_3}{\epsilon} = \frac{c_B}{\epsilon}$ can be found from (24) by setting $c_A = u_2$, $c_B = u_3$, and the exponent equal to

$$\frac{1}{\epsilon \hat{T}} = \frac{1}{\epsilon \hat{T}_+} - \frac{1}{\hat{T}_+} \frac{(u_1/\epsilon)}{1 + \epsilon(u_1/\epsilon)}.$$

The separated form (12) is found by replacing u_1 by $u_{1-} + \epsilon v_1$ and keeping only the leading order term as $\epsilon \rightarrow 0$. Doing this with the knowledge $v_{2-} = 0$ yields

$$\psi(v) = v_2 \exp[v_1 \hat{T}_+ / (\hat{T}_-)^2]$$

and

$$\eta(\epsilon) = \epsilon^2 \alpha(\epsilon) u_{3-} \exp[-1/\epsilon \hat{T}_-].$$

Since $v_1 < 0$ for u strictly between u_+ and u_- , our essential assumptions (14) and (15) follow immediately.

There is an alternate form for (13) which, being concise, may be convenient for some purposes. Assume D is diagonal and constant. For small ϵ , the quantity

$\frac{1}{\epsilon} \psi\left(\frac{u-u_-}{\epsilon}\right)$ is important only for u near u_- , where we have seen that $v = aV$, so that

$$\frac{u-u_-}{\epsilon} = aV = \frac{u_1 - u_{1-}}{\epsilon} - \frac{V}{V_1}.$$

With this substitution made in the argument of ψ in (13), the first component of that equation becomes

$$-cT_z = D_1 T_{zz} + \frac{1}{\epsilon} \psi\left(\frac{T-T_-}{\epsilon} - \frac{V}{V_1}\right) K_1.$$

The last term can in some sense be approximated by $\beta \delta(T-T_-)$, where

$$\beta = K_1 \int_{-\infty}^{\infty} \psi\left(\frac{sV}{V_1}\right) ds,$$

and we have

$$-cT_z = D_1 T_{zz} + \beta \delta(T-T_-). \quad (25)$$

This form may be immediately generalized to higher dimensional problems for which the flame front, where $T = T_-$, is not known. If the flame sheet has curvature $\gg \epsilon$, then the meaning of the δ -function is that the square of the normal derivative of T at the flame has a jump discontinuity there equal to $2\beta/D_1$. An expression using a δ -function in the space variables was used by Margolis and Matkowsky (1981).

3. Weak shocks inducing a reaction

The fronts constructed in Secs. 1 and 2 were such that the interaction between the physics and the chemistry was of a very simple nature: in the first case, the dominant feature was a physical shock in which no chemical reactions occur, and in the second, the physics was unaffected by the chemistry. There are important situations, however, when the interaction is essential to the problem, and this section contains the asymptotic analysis of such a situation. Models of this sort were studied by Fickett (1979) and Majda (1980). The present treatment is very roughly patterned after the asymptotic development in Rosales and Majda (1980), which contains many other results as well.

The three main assumptions have to do with the functions D , f , and g .

- (1) The transport is weak (replace D by ϵD) and decoupled:

$$D = \begin{pmatrix} \hat{D} & 0 \\ 0 & \tilde{D} \end{pmatrix}. \quad (26)$$

- (2) With energy a physical variable, replacing temperature (which was a chemical variable in the preceding section), the effect of the chemical variables

on the physics is weak, in the sense that

$$\hat{f} = \hat{f}(\hat{u}, \varepsilon^2 \hat{u}).$$

(3) The activation energy is high:

$$g = g\left(\frac{\hat{u} - \hat{u}_0}{\varepsilon}, \hat{u}, \varepsilon\right)$$

for some \hat{u}_0 , with g and its derivatives $O(1)$ as $\varepsilon \rightarrow 0$.

The solutions of (2) I shall construct have the appearance of weak shocks (\hat{u} is approximately \hat{u}_0) which interact with the chemistry by means of g 's strong dependence on \hat{u} . The formal expansion is as follows, up to terms of order ε .

$$\hat{u} = \hat{u}_0 + \varepsilon v; \quad \hat{u}_0 = \text{const} = \hat{u}_+; \quad v = v_0 + \varepsilon v_1; \quad v_0 \neq 0; \quad (27)$$

$$c = c_0 + \varepsilon c_1. \quad (28)$$

The function \hat{f} can be expanded to second order in ε as follows:

$$\hat{f}(\hat{u}, \varepsilon^2 \hat{u}) = \hat{f}(\hat{u}_0, 0) + \varepsilon \hat{f}'_1 v + \varepsilon^2 \hat{f}'_2 \hat{u} + \frac{1}{2} \varepsilon^2 \hat{f}'_{11} v v, \quad (29)$$

the arguments of \hat{f}'_1 etc. always being $(\hat{u}_0, 0)$. Below, I shall omit the carets from the symbol f .

Substituting (26-29) into the physical part of (2) yields (with $' = d/dz$)

$$-c v' + f_1 v' + \varepsilon [f_2 \hat{u}' + \frac{1}{2} (f_{11} v v)'] = \varepsilon (\hat{D} v')'. \quad (30)$$

The terms of $O(1)$ are

$$-c_0 v_0' + f_1 v_0' = 0.$$

This is a homogeneous system of algebraic equations for the variables v_0' . Since a nontrivial solution is desired, it is necessary to choose c_0 to be an eigenvalue of the matrix $f_1(u_0, 0)$. In the standard case, this is the sound speed.

It also follows that v_0' is a multiple of the associated eigenvector Ψ . Therefore the same is true of v_0 itself:

$$v_0(z) = \sigma(z) \Psi.$$

The terms of $O(\varepsilon)$ in (30) are

$$\begin{aligned} -c_0 v_1' + f_1 v_1' &= c_1 v_0' - f_2 \hat{u}' - \frac{1}{2} (f_{11} v_0 v_0)' + (\hat{D} v_0')' \\ &= c_1 \sigma' \Psi - f_2 \hat{u}' - A \sigma \sigma' + (\sigma' \hat{D} \Psi)', \end{aligned}$$

where $A = f_{11} \Psi \Psi$. For a solution v_1 to exist, it is necessary that the right hand side be orthogonal to the nullvector Ψ^* of the adjoint operator $f_1^* - c_0 I$. We can assume that there is only one such nullvector, and may normalize it so that $\Psi^* \Psi = 1$; then the condition is

$$c_1 \sigma' - V \cdot \hat{u}' - B \sigma \sigma' + (\bar{D} \sigma')' = 0, \quad (31)$$

where $B = A \cdot \Psi$, $\bar{D} = \hat{D} \Psi \cdot \Psi^*$, and $V = \Psi^* f_2$, the factor f_2 being the Jacobian matrix of f with respect to its second argument.

In this way, the set of physical variables has been reduced to a single variable σ . Let us pass on to the chemical portion of (2):

$$-c\tilde{u}' + \tilde{f}(\hat{u}, \tilde{u})' = \varepsilon(\tilde{D}u')' + g(v, \tilde{u}).$$

Terms of $O(1)$:

$$-c_0 \tilde{u}' + \tilde{f}(\hat{u}_0, \tilde{u})' = g(\sigma\psi, \tilde{u}).$$

In the (typical) case that $\tilde{f} = F(\hat{u})\tilde{u}$, this becomes

$$\tilde{u}' = h(\sigma, \tilde{u}) \quad (32)$$

where

$$h = \frac{g(\sigma\psi, \tilde{u})}{-c_0 + F(\hat{u}_0)}.$$

The problem has been reduced to a single wave front equation (31) for a physical variable σ , coupled with an ordinary differential equation (32) for the chemical variables \tilde{u} .

From (26), the boundary conditions at $+\infty$ are $\sigma(\infty) = 0$, $\tilde{u}(\infty) = \tilde{u}_+$, where $g(0, \tilde{u}_+) = 0$. The Rankine-Hugoniot relations (4) and the relations (5) provide necessary conditions on c_1 and σ_-, u_- for a solution to exist:

$$c_1 \sigma_- - V \cdot (u_- - u_+) - \frac{1}{2} B \sigma_-^2 = 0,$$

$$g(\sigma_- \psi, \tilde{u}_-) = 0.$$

If the latter equation could be solved for \tilde{u}_- as a function of σ_- , then the first equation might provide a one-parameter family of possible end states σ_- , c_1 being the parameter in some range of values. The existence of a solution of (31), (32) with these boundary conditions is a harder question; if \tilde{u} is one-dimensional, this problem has essentially completely been solved by Majda (1981). Computations have also been performed (Rosales-Majda 1980).

In summary, these fronts consist of weak shocks coupled with highly activated chemical reactions, with weak transport. The velocity comes out to be a perturbation of sound speed. The mathematical problem reduces to a boundary value problem for a scalar front equation coupled with an ordinary differential equation for the chemical variables.

4. Sharp Front Asymptotics for Chemical and Biological Problems

The use of (1) in connection with biology and chemical reactor theory is usually characterized by the lack of a convection term f , and all variables being chemical. In Sec. 2 of the talk, I spoke of a class of problems with no convection for which asymptotic analysis produces a front with a sharp corner at one point. There is another wide class of models, involving reaction functions g of "sigmoidal" type, whose solutions are characterized by the appearance of sharp transitions rather than sharp corners. These solutions may be genuine wave fronts in the sense used here, or dynamical structures which locally appear to be composed of sharp fronts, but do not conform globally to the strict definition. These sigmoidal models have been used in many contexts. A very few of the references are in connection with:

impulse propagation along nerve axons (Casten et al (1975));
other waves in excitable tissue (Ostrovskii and Yakhno (1975), Keener (1979));

chemical waves and patterns (Tyson and Fife (1980));
 geographic distribution of populations (Mimura and Murray (1978), Mimura and Nishiura (1978));
 patterns in developmental biology (Mimura and Nishiura (1978));
 more abstract self-organization models (Fife (1976)).

Here I will go through one example to illustrate the construction of a genuine wave front composed of a sharp transition separated by two regions of slower variation. The example consists of two equations, with the components $g_1(u)$, $g_2(u)$ vanishing on the curves indicated, and positive and negative in the regions indicated:

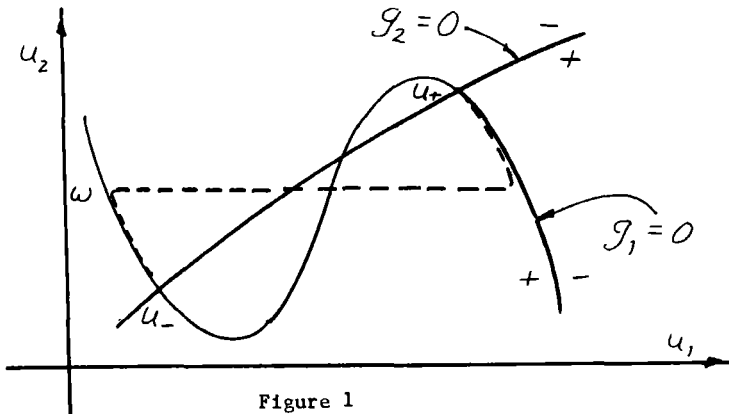


Figure 1

The component g_1 is assumed to be large (making u_1 a "fast" variable), so we replace it by $\frac{1}{\epsilon} g_1$, $\epsilon \ll 1$. The diffusion matrix is $D = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \end{pmatrix}$, meaning that u_1 diffuses more slowly than u_2 . The front equation (3) therefore takes the form

$$\epsilon u_1' + \epsilon u_1'' + \frac{1}{\epsilon} g_1(u) = 0 \quad (33a)$$

$$\epsilon u_2' + u_2'' + g_2(u) = 0. \quad (33b)$$

There are three rest states; I shall show how to construct a front passing between two of them, indicated by u_- and u_+ in the figure. Eq'n (33a) and the fact that $\epsilon \ll 1$ would suggest that $g_1(u(z)) \equiv 0$ unless u_1' or u_1'' is large. For a first try, therefore, one might attempt to fashion a front whose image in the phase plane (Fig. 1) coincides (approximately) with the portion of the curve between u_- and u_+ . But the profile of such a front would be such that u_2 has a minimum at a place (the bottom of the sigmoidal curve) where $g_2 > 0$. Since $u_2'' > 0$ there, (33b) would be violated. So this attempt is doomed to failure, and it is necessary to find a way from u_- to u_+ without following the sigmoidal curve all the way. We can follow it most of the way, however, if a horizontal jump from one branch of it to another is allowed at some (as yet unknown) value $u_2 = \omega$ as indicated in the figure. The dotted line in that figure, then, is the proper "trajectory".

Such a jump necessitates a local rescaling, so that (33) will still be satisfied. The front profile experiences a sudden transition in the u_1 component, but not the u_2 . As in the examples in Secs. 1 and 2, I shall construct the front in three parts:

- (a) For $z < 0$, (33) is to be satisfied with $\epsilon = 0$. This means $g_1(u) =$

0, and since $u(-\infty) = u_-$, the relation $u_1 = h_-(u_2)$, $u_2 < \omega$, holds, h_- being defined as the function whose graph is the left hand branch of the sigmoidal curve. Now (33b) implies

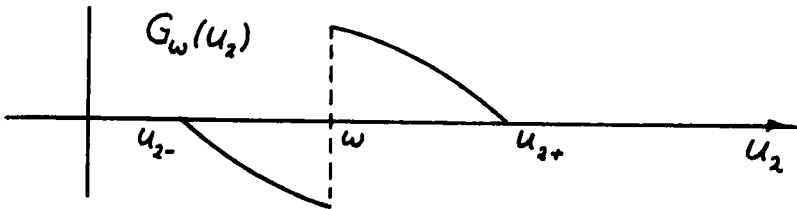
$$cu_2' + u_2'' + G_\omega(u_2) = 0 \quad (34)$$

where $G_\omega(u_2) = g_2(h_-(u_2), u_2)$ for $u_2 < \omega$.

(b) For $z > 0$, the analogous construction is to take place on the right hand branch $u_1 = h_+(u_2)$. We can still use (34) by just completing the definition:

$$G_\omega(u_2) = g_2(h_+(u_2), u_2) \text{ for } u_2 > \omega.$$

These two parts, taken together, constitute the "outer solution". The function $G_\omega(u_2)$ has the appearance:



It turns out that (34), with this type of nonlinearity, such that G_ω is negative near one rest state on the left, positive near the other, and has only one change of sign in the middle, has a unique (up to translation) solution $u(z; \omega)$ and a unique velocity

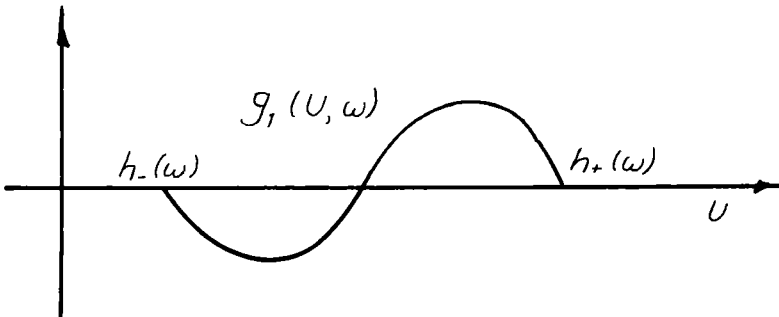
$$c = C_1(\omega). \quad (35)$$

Of course I have shown these things as depending on ω , which is not yet known. Eq. (34) is actually the equation of a scalar wave front without convection; most things are known about them, as I shall bring out in the next section.

(c) The internal structure of the shock-like transition near $z = 0$ is found by rescaling: $\zeta = \frac{z}{c}$, $u_1(z) = U(\zeta)$. This leads to

$$cU_\zeta + U_{\zeta\zeta} + g_1(U, \omega) = 0. \quad (36)$$

Here I have replaced u_2 by ω , because it stays approximately equal to that value during the transition. To match with the outer solution, U necessarily satisfies $U(\pm\infty) = h_\pm(\omega)$. By our assumption, the nonlinearity has the appearance



This has the essential properties of the previous figure, so, again, there is a unique front profile $U(\zeta; \omega)$ with a unique

$$c = c_2(\omega) \quad (37)$$

We now have to hope that (35) and (37) can be solved for ω and c . If this is the case, the formal construction is finished.

One further comment about sigmoidal reaction functions is in order. The above construction has a sharp local front imbedded in a global front. The properties of the local one are only governed by the sigmoidal component g_1 , whereas the global one depends also on g_2 . In applications, there are cases in which g_2 is such that no global front exists. Solutions with local fronts still exist, however; they are just not steady state solutions. A recent example is the analysis by Tyson and Fife (1980) of target patterns for the Belousov-Zhabotinsky reagent. Our model is based on the "Oregonator" kinetics, which has the features discussed above. Crudely speaking, what our analysis produces is a succession of local fronts, alternately of upjump and downjump types, traveling outward in both directions from the origin.

Other examples lie in the realm of signal (mainly pulse) propagation along a nerve axon. The vanguard of this approach was the work of Casten, Cohen, and Lagerstrom (1975) dealing with the FitzHugh-Nagumo equations.

II. MORE RIGOROUS RESULTS FOR SIMPLER CASES

5. A Single Equation for a Chemical Variable; No Convection.

With primes denoting differentiation with respect to z , the front equation is

$$cu' + u'' + g(u) = 0, \quad u(\pm\infty) = u_{\pm}, \quad g(u_{\pm}) = 0. \quad (38)$$

There are three types of functions g which are of most interest in connection with propagating fronts, and I shall briefly discuss the main known results for each. In the following, the function g is assumed to be smooth, although the various results probably hold for nonsmooth ones, such as the discontinuous g discussed in the last section. For definiteness, I shall take $u_- > u_+$.

$$(a) \quad g > 0 \quad \text{for} \quad u_+ < u < u_-$$

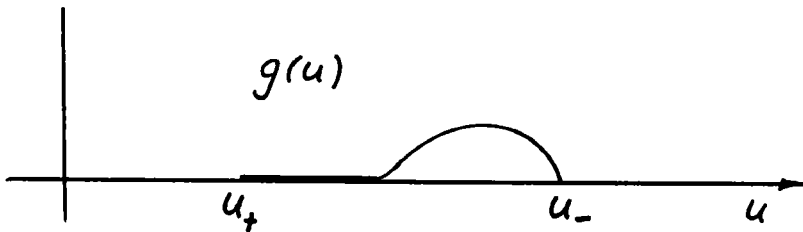
(See Kolmogorov et al (1937), Uchiyama (1978).) In this case, there is a positive minimal velocity c^* such that a traveling front exists for every speed $c > c^*$. These fronts are stable only to a very limited class of perturbations, including those of compact support (Sattinger 1978).

These fronts have been used in connection with population genetics (Fisher (1937), Aronson and Weinberger (1975)), flame theory, and chemical patterns (Tyson and Fife (1980)). In the first two applications, the front with minimal speed has proven to be the most relevant; in the third applications, however, fronts with higher velocities are essential to the construction, and appear in the guise of "phase fronts".

If $g < 0$ for $u_+ < u < u_-$, analogous results hold, except that all velocities are now negative.

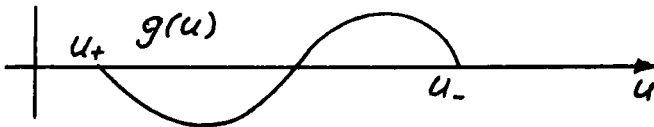
$$(b) \quad g > 0 \quad \text{for} \quad u_+ < u < u_-, \quad \text{while}$$

$$g \equiv 0 \quad \text{for} \quad u \text{ near } u_+, \quad \text{and} \quad g > 0 \quad \text{near } u_-:$$



In this case, the velocity c , and the front's profile, are both unique (Kanel' 1962). To my knowledge, the stability of this front has not been fully investigated. The main relevance of this case is in flame theory. In fact, the final model (25) derived in Sec. 2 is a limiting case, with the "hump" in the above figure becoming a δ -function.

(c) $g'(u_+) < 0$, $g'(u_-) < 0$ (the bistable case):



If there is only one intermediate zero (and in certain other cases as well), there will again be a unique front with velocity c . The velocity can be positive, negative, or zero. This front is the most stable of all, global stability having been proved in (Fife and McLeod (1977)).

These fronts have relevance to the study of nerve signal propagation, chemical patterns, population fronts, and nonlinear transmission lines.

Many of the results on scalar population fronts have been generalized by Weinberger to evolution problems of the form

$$u_{n+1} = G[u_n],$$

where the subscript n denotes discrete time, and u_n is a scalar function of either a continuous space variable ranging over all space, or a discrete space variable ranging over an infinite equally-spaced grid. In population genetics contexts (for which the model was devised), the grid points would represent colonies, and the dynamics, given by the operator G , would be given by rules of breeding, migration, and natural selection. Solutions of

$$u_t = u_{xx} + g(u)$$

may in many cases be fit into Weinberger's scheme (Weinberger 1981).

He proves that under natural conditions, fronts exist, and disturbances have, in a certain sense, a characteristic propagation speed.

6. A Single Equation for a Physical Variable:

$$-cu' + (f(u))' = u'', \quad u(\pm\infty) = u_{\pm}. \quad (39)$$

Here the Rankine-Hugoniot condition is a scalar equation:

$$c\Delta u = \Delta f(u).$$

The question of the existence of fronts is a trivial question to answer, so we proceed to the stability of existing fronts. A strong stability result, with

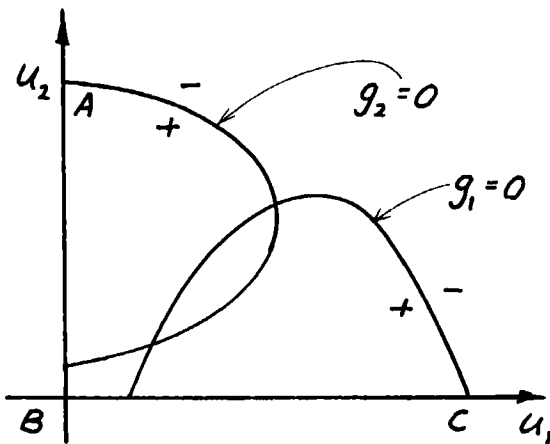
perturbations restricted to approach 0 as $z \rightarrow \infty$, was proved by Il'in and Oleinik (1960). Osher and Ralston (1981) have recently proved such a stability result when the term on the right of (39) is $(D(u)u')'$.

7. A Pair of Equations for Chemical Variables:

$$-cu' = Du'' + g(u), \quad u = (u_1, u_2),$$

positive and diagonal.

(a) Conley and Gardiner (1981) have considered the case when the nullclines of the g_1 are as follows:



and the signs of the functions g_i are appropriate for the dynamics of competing species (u_i being the densities of the two species).

When there is none of species 1 present, a scalar function of bistable type exists connecting points A and B; there is also one connecting B and C when $u_2 = 0$. Under certain conditions, Conley and Gardner proved the existence of fronts from A to C, using Conley's connection index. Gardner (1980) had previously proved the existence of fronts under certain other conditions, using Leray-Schauder theory. Stability has not been investigated.

(b) Again competing species, but with the existence of a stable coexistent rest state assumed ($u_1 \neq 0$, $u_2 \neq 0$), and with the origin unstable. Tang and Fife (1979) proved the existence of a family of fronts connecting the origin to the coexistent state, using 4-dimensional phase space analysis.

(c) A three-species model of migration with (weak) natural selection has been studied by Conley and Fife (1981). The existence of fronts was proved using Conley's index.

(d) A two-equation model proposed by Murray in connection with chemical wave problems has been investigated by Klassen and Troy (1979), who proved the existence of fronts, together with convergence results.

(e) S. S. Lin (1980) studied a model for flame theory, involving a single reaction $A + B$ with constant pressure. The dependence of the density $\rho = \rho(T)$ on temperature is not assumed to be weak. There results a system

$$U_t = (D(U)U_x)_x + \phi(U)K, \quad U = (T, Z), \quad K \in \mathbb{R}^2,$$

with $\phi(U) \equiv 0$ for T near T_+ (the temperature of the unburned gas), and for $Z = 0$. The matrix D is not a constant. The existence and uniqueness of a front was proved, using Leray-Schauder theory.

8. Conservation Laws with "Viscosity".

$$-cu' + (f(u))' = (D(u)u')', \quad u(\pm\infty) = u_{\pm}. \quad (40)$$

The most basic problems are the following:

(a) Find the set of solutions (u_+, u_-, c) of the Rankine-Hugoniot equations.

(b) Given a Rankine-Hugoniot triple (u, u_-, c) , find a solution (if it exists) of (40).

As mentioned before, these problems are trivial in the scalar case, but far from trivial for systems. Problem (b) is easily reduced to finding a trajectory between two critical points of an autonomous system of ordinary differential equations, the order equalling the number of equations in (40).

In the case of the standard equations of gas dynamics, there are three equations for the three densities of mass, momentum, and energy. Then Prob. (a) is classical. The easiest matrix D in this case is that representing viscosity and thermal conductivity. With this D (not necessarily constant) an existence theorem for Prob. (b) was proved by Gilbarg (1951).

Results on Prob. (b) for more general systems in which f satisfies strict hyperbolicity and genuine nonlinearity conditions have been obtained mainly by Conley and Smoller (1970) and (most recently) by Keyfitz (1982).

I am indebted to Basil Nicolaenko and Wildon Fickett for stimulating conversations. Support from the National Science Foundation and Los Alamos National Labs is gratefully acknowledged.

REFERENCES

- [1] Aronson, D. G. and Weinberger, H. F., Nonlinear diffusion in population genetics, combustion and nerve propagation, in: Proceedings of the Tulane Program in Partial Differential Equations and Related Topics, Lecture Notes in Mathematics, No. 446 (Springer-Berlin, 1975).

- [2] Bush, W. R. and Fendell, F. E., Asymptotic analysis of laminar flame propagation for general Lewis numbers, *Combustion Science and Technology* 1 (1970) 421-428.
- [3] Casten, R., Cohen, H. and Lagerstrom, P., Perturbation analysis of an approximation to Hodgkin-Huxley theory, *Quart. Appl. Math.* 32 (1975) 365-402.
- [4] Conley, C. and Fife, P., Critical manifolds, travelling waves and an example from population genetics, *J. Math. Biology*, to appear.
- [5] Conley, C. and Gardner, R., An application of the generalized Morse index to travelling wave solutions of a competitive reaction-diffusion model, *Math. Res. Center TSR* 2144 (1981).
- [6] Conley, C. and Smoller, J. A., Viscosity matrices for two-dimensional nonlinear hyperbolic systems, *Comm. Pure Appl. Math.* 23 (1970) 867-884.
- [7] Fickett, W., Detonation in miniature, *Am. J. Phys.* 47 (1979) 1050-1059.
- [8] Fife, P. C., Pattern formation in reacting and diffusing systems, *J. Chem. Phys.* 64 (1976), 854-864.
- [9] Fife, P. C. and McLeod, J. B., The approach of solutions of nonlinear diffusion equations to travelling front solutions, *Arch. Rational Mech. Anal.* 65 (1977), 335-361.
- [10] Gardner, R. A., Large amplitude patterns for two competing species, *Math. Res. Center TSR*. (1980).
- [11] Gilbarg, D., The existence and limit behavior of the one-dimensional shock layer, *Amer. J. Math.* 73 (1951) 256-274.
- [12] Kanel', Ya. I., On the stabilization of solutions of the Cauchy problem for the equations arising in the theory of combustion, *Mat. Sbornik* 59 (1962) 245-288.
- [13] Keener, J. P., Waves in excitable media, *SIAM J. Appl. Math.* 39 (1980) 528-548.
- [14] Keyfitz, B., Bounds for viscosity profiles for 2×2 systems of conservation laws, *Rocky Mountain J. of Math.*, to appear.
- [15] Klaasen, G. and Troy, W., The asymptotic behavior of solutions of a system of reaction-diffusion equations which models the Belousov-Zhabotinskii chemical reaction, *J. Differential Eq'ns*, in press.
- [16] Lin, S. S., Theoretical study of a reaction diffusion model for flame propagation in a gas, Ph.D. Thesis, University of California, Berkeley, *Math. Res. Center TSR* (1980).
- [17] Linan, A., A theoretical analysis of premixed flame propagation with an isothermal chain reaction, *Tech. Report, Inst. Nac. Tec. Aeroespacial "Esteban Terradas"*, Madrid (1971).
- [18] Majda, A., A qualitative model for dynamic combustion, *SIAM J. Appl. Math.*, to appear.
- [19] Margolis, S. B. and Matkowsky, M. J., Flame propagation with multiple fuels, *SIAM J. Appl. Math.*, to appear.

- [20] Mimura, M. and Murray, J. D., On a planktonic prey-predator model which exhibits patchiness, *J. Theoret. Biology* 75 (1979) 249-262.
- [21] Mimura, M. and Nishiura, Y., Spatial patterns for interaction-diffusion equations in biology, *Proc. Symp. on Mathematical Topics in Biology at RIMS, Kyoto University* (1978).
- [22] Osher, S. and Ralston, J., L_1 stability of traveling waves with applications to convection porous media flow, preprint (1981).
- [23] Ostrovskii, L. A. and Yakhno, V. G., The formation of pulses in an excitable medium, *Biofizika* 20 (1975) 489-493.
- [24] Rosales, R. R. and Majda, A., Weakly nonlinear detonation waves, *SIAM J. Appl. Math.*, to appear.
- [25] Sattinger, D., Weighted norms for the stability of traveling waves, *J. Differential Equations*, 25 (1978) 130-144.
- [26] Tang, Min Ming and Fife, P. C., Propagating fronts for competing species equation with diffusion, *Arch. Rational Mech. Anal.*, 73 (1979) 69-77.
- [27] Tyson, J. and Fife, P. C., Target patterns in a realistic model of the Belousov-Zhabotinskii Reaction, *J. Chem. Physics* 73 (1980) 2224-2237.
- [28] Uchiyama, K., The behavior of solutions of some non-linear diffusion equations for large time, *J. Math. Kyoto Univ.* 18 (1978) 453-508.
- [29] Williams, F. A., *Combustion Theory*, (Addison-Wesley, Reading, MA. 1965).
- [30] Weinberger, H.F., Long-time behavior of a class of biological models, *SIAM J. Math. Anal.*, to appear.

This Page Intentionally Left Blank

PART IV
Nonlinear Phenomena
in Fluids and Plasmas

This Page Intentionally Left Blank

REGULARITY RESULTS FOR THE EQUATIONS OF INCOMPRESSIBLE
FLUIDS MECHANIC AT THE BRINK OF TURBULENCE

Claude BARDOS
Département de Mathématiques
Université of Paris 13
Avenue J.B. Clément 93430 VILLETANEUSE

This work is divided in two parts firsts we give the existing regularity results concerning the equations of incompressible fluids mechanic. These results are valid in a limited range of applications. It is conjectured that this limitation is due to the appearance of turbulence. Therefore in a second part we describe some recent numerical results obtained in situations where the abstract theorems are no more valid; these computations give evidence of persistence of regularity for the two dimensional M.H.D. equations and of appearance of Turbulence after a finite time for the 3 dimensional Euler Equation.

I. INTRODUCTION

The purpose of this lecture, on one hand, is to describe the classical results concerning the existence of a smooth solution for the equations of incompressible fluids mechanic. These results have a rather limited range of applications and it is conjectured that they cease to be valid when turbulence appears; therefore, on the other hand, I will try to describe some recent numerical results concerning the behaviour of the fluid at the brink of turbulence (when the theoretical results are no more available). These numerical experiments would hopefully give some light on the process; however they require a very sophisticated material both technical (size of the programs involved) and theoretical (use for instance of Euler Transform or Padé approximants) and in the future their interpretation may be subject to important changes.

II. THEORETICAL REGULARITY RESULTS

1. The basic equations and the fundamental regularity results

The function $u(x,t)$ will denote the velocity; it is a vector valued function defined in $\mathbb{R}_x^n \times \mathbb{R}_t$ ($n = 1, 2, 3$) with value in \mathbb{R}^n ; for numerical purpose, and also to have at our disposal a problem in a bounded domain with no boundaries we will also consider the periodic case where $x \in T^n$ and u takes its values in T^n . Therefore X will denote either \mathbb{R}_x^n or T^n , in the first case we will be concerned with functions u that decay at infinity (in some way) and in the second, with periodic functions.

The incompressible viscid Navier Stokes equation is then :

$$\frac{\partial u}{\partial t} + u \cdot \nabla u = -\nabla p + \nu \Delta u, \quad \nabla \cdot u = 0 \quad (1)$$

Incompressible refers to the relation $\nabla \cdot u = 0$ and this is equivalent to the following fact : the mapping ψ defined by

$$\dot{x}(t) = u(x(t), t), \quad x(0) = x, \quad \psi(x, t) = x(t) \quad (2)$$

is measure preserving. On the other hand the positive number ν denotes the viscosity, and the relative strenght of the non linear term and of the viscous term is measured by the dimensionless Reynold number

$$R = \frac{\text{Typical lenght} \times \text{Typical velocity}}{\text{viscosity}} = \frac{|\nu \cdot \nabla \nu|}{|\nu \cdot \Delta \nu|}$$

For $\nu = 0$ the equation (1) is called the Euler equation.

For a conductive fluid, when the variation of the electric field is neglected the motion is governed by two set of equations involving the velocity u and the magnetic field b :

$$\frac{\partial u}{\partial t} + u \cdot \nabla u = -\nabla p + \nu \Delta u + b \nabla b, \quad \nabla \cdot u = 0 \quad (\nu \geq 0) \quad (3)$$

$$\frac{\partial b}{\partial t} + u \cdot \nabla b = b \nabla u + \eta \Delta b, \quad \nabla \cdot b = 0 \quad (\eta \geq 0) \quad (4)$$

Some over simplified models of these equations can be given in one space variable

they are the following :

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = 0 \quad (5)$$

(Burger Hopf equation) or

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = b \frac{\partial b}{\partial x} \quad (6)$$

$$\frac{\partial b}{\partial t} + u \frac{\partial b}{\partial x} = b \frac{\partial u}{\partial x} \quad (7)$$

2. Regularity results for the Navier Stokes Equation

It is known that for $n = 2$ the equation (1) with $\nu > 0$ and the equations (3), (4) with $\nu > 0$ and $\eta > 0$ admit a smooth solution defined for all positive times (Leray [19], Lions [21], Ladyshenskaya [18] for the Navier Stokes Equation and C. Sulem [28] for the M.H.D. equation); furthermore the solution is analytic in t (for $t > 0$) more precisely it can be extended in a complex neighbourhood of \mathbb{R}^2 (or T^2 for periodic functions) of the following shape :

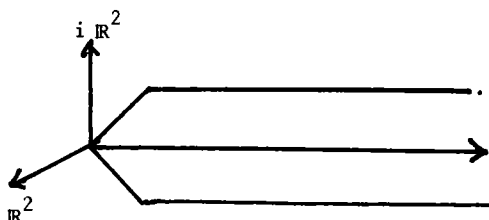


Figure 1

(cf. Kato Fujita [14] or Foias Prodi [8]).

The proofs of these results use basically the properties of the Laplace operator to balance (via Sobolev Theorem) the effect of the non linear terms.

In dimension 3 similar results are proven only for a viscosity large enough (corresponding to a Reynold number of the order of one). For higher Reynolds numbers only the existence of a weak solution is proven. However one can show that the size of the domain (in x and t) where this solution may fail to be analytic is very small. Several results in this direction have been obtained starting with Leray [17] and then improved by Scheffer [27], Foias Teman [9] and Caffarelli and Nirenberg [5]. The result of [5] is the following : the Hausdorff dimension in $\mathbb{R}_x^3 \times \mathbb{R}_t$ of the set of points where the solution fails to be smooth

is at most 1. The proof it self is very interesting because it rests mainly on dimension analysis and therefore it is very close of the heuristic argument of Kolmogorov [16] concerning the spectra of the Turbulence .

3. Euler Equation in 3 dimensions

For the Euler equation there is no proof of the existence of a solution (even a weak one) for all time. The only result concerns the existence (and uniqueness) of a smooth solution for a finite time $|t| < T^*$ depending on the size of the initial data and mainly on the size of the initial vorticity (Lichtenstein [20], Ebin and Marsden [7]), and the following facts have been observed :

(i) If the initial data belongs to C^k then as long as the vorticity remains bounded the solution will remain in the same space (Ebin and Marsden [7] or Foias Frisch and Teman [10]).

(ii) If the initial data is analytic then the solution remains analytic as long as the vorticity is bounded. However the complex domain where the solution is analytic may shrink and finally collapse when the vorticity cease to be bounded.

This gives, for the domain of analyticity the following picture. (Bardos and Benachour [3]).

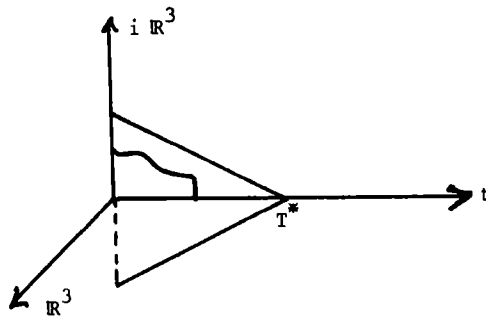


Figure 2

The proof of all these results are done with the system :

$$\frac{\partial \omega}{\partial t} + u \nabla \omega = \omega \nabla u ; \quad \nabla u = 0 ; \quad \nabla \wedge u = \omega \quad (8)$$

and the introduction of the characteristics curves of the motion

$$\dot{x}(t) = u(x(t), t) , \quad x(T) = x \quad (9)$$

For (ii) an extension of these curves to the complex domain is necessary and one has to control their distance from \mathbb{R}^2 in term of the Holder norm of the vorticity.

4. Euler Equation in two dimension

In two dimension the vorticity $\omega = \nabla \wedge u$ is orthogonal to the plane (x_1, x_2) and therefore the term $\omega \nabla u$ cancels. The vorticity is conserved along any trajectory of the fluid particle and its norm in any L^p space ($1 \leq p \leq \infty$) is constant. From this observation one deduces by a compactness argument the existence of a weak solution for any initial data $u(0, \cdot)$ of finite energy with $\nabla \wedge u(0, \cdot) = \omega(0, \cdot)$ bounded in L^p . When $p = +\infty$ one can also prove the uniqueness of the solution (Youdovitch [33], Bardos [2]). The case $p = 1$ is slightly more difficult because the unit ball of L^1 is not weakly closed.

It has been noticed by Wolibner [32] and Holder [13] that from the uniform boundedness of the curl one cannot directly deduce the regularity of the solution. One has to introduce an other argument concerning the pair dispersion. The problem is to control the distance between two fluid particles $x(t)$ and $y(t)$ knowing that the velocity field $u(x, t)$ is divergence free and has bounded curl. Due to the divergence logarithmic singularity of the Green function, for the Laplace equation $\Delta \psi = \omega$, one has the estimate

$$\left| \frac{\partial^2 \psi}{\partial x_i \partial x_j}(x) - \frac{\partial^2 \psi}{\partial x_i \partial x_j}(y) \right| \leq C |\omega|_\infty |x-y| \operatorname{Log} \frac{1}{|x-y|} \quad (1)$$

With $u = \nabla \wedge \psi = \left(\frac{\partial \psi}{\partial y}, -\frac{\partial \psi}{\partial x} \right)$ one deduced from (10) the following a priori estimate :

$$|\dot{\rho}(t)| = \left| \frac{d}{dt} |x(t) - y(t)| \right| \leq C |\omega|_\infty \rho \operatorname{Log} \frac{1}{\rho} \quad (11)$$

Solving the ordinary differential equation $z' = \pm C |\omega|_\infty z \operatorname{Log} \frac{1}{z}$ and using a comparison theorem, one obtains the estimate :

(1) For sake of simplicity we consider a periodic domain of diameter 1.

$$\frac{\exp(C|\omega|_{\infty}t)}{\rho(0)} \leq \rho(t) \leq \frac{\exp(-C|\omega|_{\infty}t)}{\rho(0)} \quad (12)$$

from which the Holder regularity of u can be deduced.

Similar methods can be extended to complex domain and one obtains (Bardos Benachour [3]) that if the initial data is analytic in a strip

$$\{x + iy = (x_1 + iy_1, x_2 + iy_2) \mid |y| < \eta\},$$

the solution will be analytic in a strip of the following form :

$$\{x + iy ; |y| < \eta \exp(-C e^{Kt}) / \exp(-C)\}$$

5. Kelvin Helmholtz instability in two dimensions

The Euler equation can be written in conservative form :

$$\frac{\partial}{\partial t} u_j + \frac{\partial}{\partial x_i} (u_i u_j) = - \frac{\partial p}{\partial x_j}, \quad i, j = 1, 2 ; \quad \nabla \cdot u = 0 \quad (13)$$

therefore one can try to solve the initial value problem for some class of discontinuous initial data. In particular we will assume that $\omega_0 = \nabla \wedge u_0$ is a smooth density supported by the smooth curve $\Gamma = \{(x, y(x, t)) \mid x \in \mathbb{R}\}$ and that u_0 is of finite energy.

Let $\rho \in \mathcal{D}(\mathbb{R}^2)$ be a positive function satisfying the relation $\int \rho(x) dx = 1$. We introduce the solution u^ε of the Euler equation, with initial data given by

$$u^\varepsilon(\cdot, 0) = \varepsilon^{-2} \rho\left(\frac{\cdot}{\varepsilon}\right) * u_0(\cdot) \quad (14)$$

The energy of $u^\varepsilon(\cdot, 0)$ is uniformly bounded and we have, using the conservation of the vorticity, the relation :

$$\begin{aligned} \int_{\mathbb{R}^2} |\nabla \wedge u^\varepsilon(x, t)| dx &= \int_{\mathbb{R}^2} |\nabla \wedge u^\varepsilon(x, 0)| dx \leq \int_{\mathbb{R}^2} dx \varepsilon^{-2} \rho\left(\frac{x-\sigma}{\varepsilon}\right) \int_{\Gamma} |\omega_0(\sigma)| d\sigma \\ &\leq \int_{\Gamma} |\omega_0(\sigma)| d\sigma \end{aligned} \quad (15)$$

Therefore the curl of u_0^ε remains uniformly bounded (with respect to ε) in $L^1(\mathbb{R}^2)$. Since u_ε is bounded in $L^\infty(\mathbb{R}_t ; (L^2(\mathbb{R}^2))^2)$ one can extract a sub family

which will converge (with ε going to zero) in $L^\infty(\mathbb{R}_t; (L^2(\mathbb{R}^2))^2)$ weak star to a function u ; u is a solution of the equation

$$\frac{\partial u_j}{\partial t} + \frac{\partial}{\partial x_i} (\eta_{ij}) = - \frac{\partial p}{\partial x_i} \quad (16)$$

with $\eta_{ij} = \lim_{\varepsilon \rightarrow 0} u_i^\varepsilon \cdot u_j^\varepsilon$ in $\mathcal{D}'(\mathbb{R}_t \times \mathbb{R}_x^2)$.

To obtain a weak solution one has to prove the relation :

$$\lim_{\varepsilon \rightarrow 0} u_i^\varepsilon u_j^\varepsilon = u_i u_j \quad (17)$$

As it is usually the case one proves (17) using a compactness argument and for this compactness argument we have at our disposal the estimate (15) which is the essential step. Therefore we obtain (see C. Sulem, P.L. Sulem, C. Bardos and U. Frisch [30] for the details) the following.

Theorem : Assume that u_0 belongs to the space $(L^2(\mathbb{R}^2))^2$ and that we have

$$\nabla \wedge u_0 = \omega_0(\sigma) \otimes \delta_\Gamma$$

where Γ denotes a smooth curve and $\omega_0(\sigma)$ a smooth density satisfying the estimate :

$$\int_\Gamma |\omega_0(\sigma)| d\sigma < +\infty$$

Then the Euler equation in $\mathbb{R}_t \times \mathbb{R}^2$:

$$\frac{\partial u}{\partial t} + u \nabla u = - \nabla p, \quad \nabla \cdot u = 0$$

with initial data $u(\cdot, 0) = u_0(\cdot)$ has a weak solution belonging to the space $L^\infty(\mathbb{R}; (L^2(\mathbb{R}^2))^2)$.

It is important to notice that $\nabla \wedge u$ will not belong to $L^1(\mathbb{R}^2)$. This is related to the fact that $L^1(\mathbb{R}^2)$ is not reflexive. The introduction of the estimate in $L^1(\mathbb{R}^2)$ is very similar to the method used to prove the existence of a weak solution for quasilinear equation (or systems) cf. Kruckov [17] or Glimm [12]; it is also related to the introduction (cf. Temam and Strang [31]) of the space

of Bounded Deformation for plasticity problems.

The situation described above gives an example of a weak solution for the Euler Equation. As long as the set where u is singular remains a smooth curve $\Gamma(t)$, one can derive from (13) that u satisfies on $\Gamma(t)$ a Rankine Hugoniot condition ; at variance with the situation for quasilinear equations or systems we will prove (Theorem 2) the uniqueness of the solution without any entropy condition. This is related to the fact that $\Gamma(t)$ is a contact discontinuity with no loss of energy.

The question of the existence and the smoothness of the curve $\Gamma(t)$ is called the Kelvin Helmholtz instability. We will now describe this problem.

We assume that the second coordinate of the curve $\Gamma(t) = \{x(\lambda, t), y(\lambda, t) ; \lambda \in \mathbb{R}\}$ can be resolved in terms of the first :

$$y = y(x, t) \quad (18)$$

and we denote by $\Omega(x, t)$ the vortex density

$$\Omega(x, t) = \omega(x, t) \sqrt{1 + (y'_x)^2} \, dx \quad (19)$$

and by $v(x, t) = \frac{u_+(x, t) + u_-(x, t)}{2}$ the mean value of the velocity on the vortex line (u_{\pm} denoting the limit of the velocity above and below the curve $\Gamma(t)$).

From the fact that u is a weak solution of the Euler equation :

$$\frac{\partial u_j}{\partial t} + \sum_i \frac{\partial}{\partial x_i} (u_i u_j) = 0, \quad \nabla \cdot u = 0 \quad (20)$$

and from the fact that u satisfies the relation :

$$\nabla \wedge u = \omega \otimes \delta_{\Gamma(t)} \quad (21)$$

one deduces that Ω and v satisfy the system

$$\frac{\partial y}{\partial t} + \frac{\partial y}{\partial x} v_1 = v_2 \quad (22)$$

$$\frac{\partial}{\partial t} \Omega + \frac{\partial}{\partial x} (\Omega v_1) = 0 \quad (23)$$

Finally from the relations $\nabla \cdot u = 0$ and (21) we have :

$$v_1 = -\frac{1}{2\pi} \int \frac{y(x,t) - y(x',t)}{(x-x')^2 + (y(x,t) - y(x',t))^2} \Omega(x',t) dx' \quad (24)$$

$$v_2 = \frac{1}{2\pi} \int \frac{x - x'}{(x-x')^2 + (y(x,t) - y(x',t))^2} \Omega(x',t) dx' \quad (25)$$

The equations (22), (23), (24) and (25) were directly derived by Birkoff [4] using mechanical arguments.

This problem is known to be highly unstable and in fact we will show that the linearised problem may be badly posed (in $\mathcal{F}(\mathbb{R})$ for instance). We assume that for $t = 0$ the vortex line is flat namely that we have :

$$u((x_1, x_2), 0) = \begin{cases} (-A, 0) & \text{for } x_2 > 0 \\ (A, 0) & \text{for } x_2 < 0 \end{cases} \quad (26)$$

Therefore we have $\Omega(0, x) \equiv 2A$ and $y(X, 0) \equiv 0$; it is easy to verify that u given by (26) is a stationary solution of (20). Writing $\Omega = 2A + \delta\Omega$, $y = 0 + \delta y$, $v_1 = 0 + \delta v_1$, $v_2 = 0 + \delta v_2$, we linearise the equations (22), (23), (24) and (25) in the neighbourhood of this stationary state, to obtain :

$$\frac{\partial}{\partial t} \delta y = \delta v_2, \quad \frac{\partial}{\partial t} (\delta\Omega) + 2A \frac{\partial}{\partial x} (\delta v_1) = 0 \quad (27)$$

$$\delta v_1 = -\frac{2A}{2\pi} \int_{\mathbb{R}} \frac{\delta y(x,t) - \delta y(x',t)}{(x-x')^2} dx' \quad (28)$$

$$\delta v_2 = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{1}{(x-x')} \delta\Omega dx' \quad (29)$$

Taking the Fourier Transform of these equations with respect to x and denoting ξ the dual variable, we obtain :

$$\frac{\partial}{\partial t} \begin{pmatrix} \widehat{\delta\Omega}(\xi) \\ \widehat{\delta y}(\xi) \end{pmatrix} + A(\xi) \begin{pmatrix} \widehat{\delta\Omega}(\xi) \\ \widehat{\delta y}(\xi) \end{pmatrix} = 0 \quad (30)$$

In (30) $A(\xi)$ denotes the matrix given by :

$$A(\xi) = A \begin{pmatrix} 0, & 2i\xi^2 \operatorname{sign} \xi \\ \frac{\operatorname{sign} \xi}{2i}, & 0 \end{pmatrix} \quad (31)$$

The eigen values of $A(\xi)$ are $\pm \xi A$ therefore $e^{tA(\xi)}$ will introduce in the ξ mode an exponential factor $e^{\pm t \xi A}$ this implies that even for $(\delta\Omega, \delta y) \in (\mathcal{J}(\mathbb{R}))^2$ ⁽¹⁾ $e^{tA(\xi)} (\delta\Omega, \delta y)$ will not belong to $(\mathcal{J}'(\mathbb{R}))^2$ and this shows that the problem is badly posed, except for initial data with exponentially decaying Fourier Transform.

The fact that the Fourier Transform of the initial data decay exponentially implies that the initial data it self is analytic, and this means that the problem (22), (23), (24), (25) may be well posed only (even for a finite time) in the class of analytic function. Indeed we have the following.

Theorem 2 : Assume that $y_0(x)$ and $\Omega_0(x)$ are the restriction to the real axis of functions $y_0(x + i\eta)$, $\Omega_0(x + i\eta)$ which are holomorphic and bounded in a strip $S = \{x + i\eta \mid |\eta| < S\}$, and assume furthermore that $\frac{\partial y_0}{\partial x}$ satisfies the estimate :

$$\left| \operatorname{Im} \frac{\partial y_0}{\partial x} (x + i\eta) \right| < 1/2 \quad (32)$$

Then the problem (22) - (25) ; with initial data y_0, Ω_0 has, during a finite time $|t| < T^*$, a unique analytic solution.

The proof of the Theorem 2 can be found in C. Sulem, P.L. Sulem, C. Bardos and U. Frisch [30]. The main idea of the proof is the following : the problem (22) - (25) is (after elimination of v_1, v_2 , using the equations (24) and (25)) of the following form :

$$\frac{\partial}{\partial t} \begin{pmatrix} y \\ \Omega \end{pmatrix} + A \begin{pmatrix} y \\ \Omega \end{pmatrix} = 0 \quad (33)$$

Where A denotes a first order non linear pseudo differential operator. Therefore one can applies the Cauchy Kowalevski theorem in scale of spaces (here we consider scales of spaces of analytic functions, cf. Ovsjannikov [26] and Baouendi and Goulaouic [1]). The main drawback of the method is the following we have to control the quantity $(x-x')^2 + (y(x,t) - y(x',t))^2$ which appears in the deno-

⁽¹⁾ \mathcal{J} denotes the space of Schwartz functions.

minator of (24) and (25). This quantity may vanishes for y non real and for $x' \neq x$, to prevent the appearance of this type of singularity we assume that the hypothesis (22) is satisfied and we consider only time intervals during which $|\frac{\partial y}{\partial z}|$ remains bounded away from 1. Therefore it seems that the theory will not be able to handle cases where the curve Γ rolls up before breaking. However some numerical computations made by Meiron, Baker and Orszag [22] seem to indicate a loss of analyticity of the curve Γ before the formation of rolls.

6. Equations of Magneto-hydrodynamics

For the two dimension M.H.D. one can prove the existence for all times of a smooth solution for non zero fluid viscosity ν and non zero magnetic viscosity (C. Sulem [28]). However when the fluid is inviscid only local existence is proven. The situation is the same as in the case of the Euler equation in 3 dimension. In particular one can obtain a local existence result for analytic solution in a domain similar to the one given by the figure 2.

On the other hand the one dimensional model :

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = b \frac{\partial b}{\partial x}, \quad \frac{\partial b}{\partial t} + u \frac{\partial b}{\partial x} = b \frac{\partial u}{\partial x} \quad (34)$$

has been studied by C. Sulem, J.D. Fournier, U. Frisch and P.L. Sulem [29]. It is clear that if b is identically zero the system (34) reduces to the Burger equation :

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0 \quad (35)$$

and therefore, for any non increasing initial data, singularities will appear after a finite time.

This property remains true for the system (34) when $b(x,0)$ is small compared to the variation of $u(x,0)$. Namely we have.

Proposition 1 : Assume that there exist a point $a \in \mathbb{R}$ such that one has for the initial data $b_0(x)$ and $u_0(x)$:

$$b_0(a) = 0 \quad \text{and} \quad -|b'_0(a)| > u'_0(a) \quad (36)$$

let $x(t)$ be the trajectory defined by :

$$\dot{x}(t) = u(x(t), t) \quad \text{and} \quad x(0) = a \quad (37)$$

then $\frac{\partial}{\partial x} u(x(t), t) + \frac{\partial b}{\partial x}(x(t), t)$ will become infinite in a finite time.

Proof - (cf. C. Sulem, J.P. Fournier, U. Frisch and P.L. Sulem [29]).

One introduces the function z^\pm defined by $z^+ = u+b$, $z^- = u-b$, with these new functions the system (34) becomes

$$\frac{\partial z^+}{\partial t} + z^- \frac{\partial z^+}{\partial x} = 0, \quad \frac{\partial z^-}{\partial t} + z^+ \frac{\partial z^-}{\partial x} = 0 \quad (38)$$

On the curve $x(t)$ defined by (37) one has $b(x(t), t) = 0$ and therefore using the relation :

$$\dot{x}(t) = u(x(t), t) = z^+(x(t), t) = z^-(x(t), t), \quad (39)$$

one obtains

$$\frac{d}{dt} \left(\frac{\partial z^+}{\partial x}(x(t), t) \right) = \frac{d}{dt} \left(\frac{\partial z^-}{\partial x}(x(t), t) \right) = - \frac{\partial z^+}{\partial x}(x(t), t) \frac{\partial z^-}{\partial x}(x(t), t) \quad (40)$$

(40) can be solved explicitly a to give :

$$\frac{\partial z^+}{\partial x}(x(t), t) = \frac{2b'_0(a)z_0^{+'}(a)}{z_0^{+'}(a) - z_0^{-'}(a)e^{-2b'_0(a)t}} \quad (41)$$

From (41) one deduces, using (39) the proof of the proposition 1.

On the other hand it is believed and known in some physical examples that a large magnetic field may stabilize the fluid and prevent the appearance of the Turbulence ; on our one space variable model this gives the :

Proposition 2 [29] : Assume that the initial conditions u_0 and $b_0 \in C^\infty(\mathbb{R})$ satisfy the relation :

$$\inf |b_0| > \frac{1}{2} (\sup u_0 - \inf u_0) \quad (42)$$

then the solution of (34) is smooth for all times.

Proof : From the relations (38) we deduce the equation

$$\left(\frac{\partial}{\partial t} + z^- \frac{\partial}{\partial x}\right)(z^+ + z^-) = (z^- - z^+) \left(\frac{\partial z^-}{\partial x}\right) \quad (43)$$

which combined with the relation

$$\left(\frac{\partial}{\partial t} + z^- \frac{\partial}{\partial x}\right) \left(\frac{\partial z^+}{\partial x}\right) = - \frac{\partial z^-}{\partial x} \frac{\partial z^+}{\partial x} \quad (44)$$

gives

$$\left(\frac{\partial}{\partial t} + z^- \frac{\partial}{\partial x}\right) \left((z^+ - z^-) \frac{\partial z^+}{\partial x}\right) = 0 \quad (45)$$

Assuming that $z^+(x,t) - z^-(x,t) = 2b(x,t)$ never vanishes and integrating (45) from zero to t we obtain

$$\frac{\partial z^+}{\partial x}(x,t) = \frac{\partial z_0^+(a^+)}{\partial x} \frac{2b_0(a^-)}{z_0^+(a^-) - z_0^-(a^+)} \quad (46)$$

In the equation (46) a^+ and a^- denote the values at $s = 0$ of the solutions of the characteristic equations

$$\dot{x}^\pm(s) = z^\pm(x(s), s), \quad x^\pm(t) = x \quad (47)$$

Finally the relation (42) implies that $z_0^+(a^-) - z_0^-(a^+)$ is never equal to zero and therefore $\frac{\partial z^+}{\partial x}$ remains uniformly bounded. Similar result yields for $\frac{\partial z^-}{\partial x}$ and from these estimates one deduces that z^\pm are smooth for all times.

Remark 1 : A similar result can be proven for the true M.H.D. equation in \mathbb{R}^n (Bardos, Frisch, Sulem unpublished) in the following situation, one assume that $b_0 = B_0 + \tilde{b}_0$ where B_0 is a constant vector field of large enough magnitude compared to \tilde{b}_0 and u_0 and that \tilde{b}_0 and u_0 decay fast enough for $|x| \rightarrow \infty$ (eventually one can take \tilde{b}_0 and u_0 of compact support), then the M.H.D. equation

$$\frac{\partial u}{\partial t} + u \nabla u = b \nabla b - \nabla p, \quad \frac{\partial b}{\partial t} + u \nabla b = b \nabla u, \quad \nabla \cdot u = \nabla \cdot b = 0 \quad (48)$$

written on the form

$$\frac{\partial Z^{\pm}}{\partial t} + Z^{\pm} \frac{\partial Z^{\pm}}{\partial t} = -\nabla p$$

Can be viewed as small perturbation of the simple transport equation

$$\frac{\partial Z^{\pm}}{\partial t} + B_0 \cdot \nabla Z^{\pm}$$

whose solutions are given by :

$$Z^{\pm}(x, t) = Z_0^{\pm}(x \mp B_0 t)$$

and in particular for the curl of Z^{\pm} one has :

$$\frac{\partial(\nabla \wedge Z^{\pm})}{\partial t} + Z^{\pm} \cdot \nabla(\nabla \wedge Z^{\pm}) = \pi(Z^+, Z^-) \quad (49)$$

where $\pi(Z^+, Z^-)$ is a term of the following form :

$$(\pi(Z^+, Z^-))_r = \alpha_{ijkl}^r \frac{\partial Z_i^+}{\partial x_k} \frac{\partial Z_j^-}{\partial x_l}$$

where α_{ijkl}^r denotes a suitable constant.

This term is quadratic but the functions ∇Z^+ and ∇Z^- propagate roughly in opposite direction (of the order of $\mp B_0$) therefore this phenomena of propagation balances the non linearity and may prevent the appearance of singularity.

The idea of this proof is related to the proofs of regularity for non linear wave equation which rely on the dispersion property of the linear wave equation in \mathbb{R}^n (cf. Klainerman [15] or Chadam [6]) and they cannot be adapted to the case of a bounded domain or of a periodic domain.

III. NUMERICAL EXPERIMENT

In the previous section we have describe regularity results concerning the classical incompressible equations of Fluids Mechanic. We have shown that these results have a limited range of application. The purpose of these section is to give some comment on numerical computations which have been made outside this range of application, namely when turbulence may appear.

1. Magneto hydrodynamic equations in two dimension

The computation described in this section have been made by Orszag and Tang [25] for the two dimensionnal periodic M.H.D. equation, with the following initial conditions :

$$\begin{aligned} u_1(x_1, x_2, 0) &= -\sin x_2, \quad u_2(x_1, x_2, 0) = \sin x_1, \\ b(x_1, x_2, 0) &= \text{curl}(a(x_1, x_2, 0)) = \text{curl}(\cos x_2 + \frac{1}{2} \cos 2x_1) \end{aligned} \quad (50)$$

there initial conditions are called the Orszag - Tang - Vortex, figure 3 and figure 4 show these initial conditions, figure 5 and figure 6 show the behaviour of the energy spectrum which is proportional to the quantity :

$$E(\lambda) = \sum_{\substack{\vec{k} \\ |\vec{k}|=\lambda}} |\hat{u}(t, \vec{k})|^2 \quad (51)$$

Where \vec{k} denotes the wave number $\hat{u}(t, \vec{k})$ the corresponding Fourier coefficients and λ the wave length. The computation shows that $E(\lambda)$ behaves like $e^{-\delta(t)\lambda}$ which implies that $u(t, x)$ can be extended as an analytic function in the complex neighbourhood of width $\delta(t)$ of the real domain. On the figure 7 is plotted for several value of t the expression $\text{Log } \delta(t)$. This gives the estimate $\delta(t) \approx -Kt$ and seems to indicate that for any t the solution will remain analytic in a complex neighbourhood of width e^{-Kt} of the real domain. The mechanism which prevent the formation of singularities may be explained by the numerical results described on figures 8 and 9, where $\xi^\pm = \text{curl } Z^\pm$ are plotted ; these terms become large only in the neighbourhood of neutral points ; around these points the phenomena becomes quasi one dimensional and therefore the Lorentz force $\pi(Z^+, Z^-)$ introduced in the equation (49) becomes un effective even for ξ^+ and ξ^- large. Finally the oscillations of the energy spectra in the figure 6 are related to the following fact : The small scale structures are located around several distincts neutral points and this produces interferences in Fourier Spectra.

On the figure 10 are plotted three curves describing the logarithmic decrement of the energy spectra . The first one concerns the computations described above for

the M.H.D. equation, the second concerns some computations made by Frisch and Sulem for the 2 dimension incompressible Euler equation, by high resolution spectral methods on an initial data suggested by Meiron, The third one is related to the bound deduced from the complex version of the pair dispersion formula (Bardos and Benachour [3]). This picture shows that the Euler equation is "more" analytic than the M.H.D. equation and that the theoretical results of [3] should be sensibly improved.

Orszag - Tang - Vortex
Stream function contour
 $u = \text{curl} \psi$

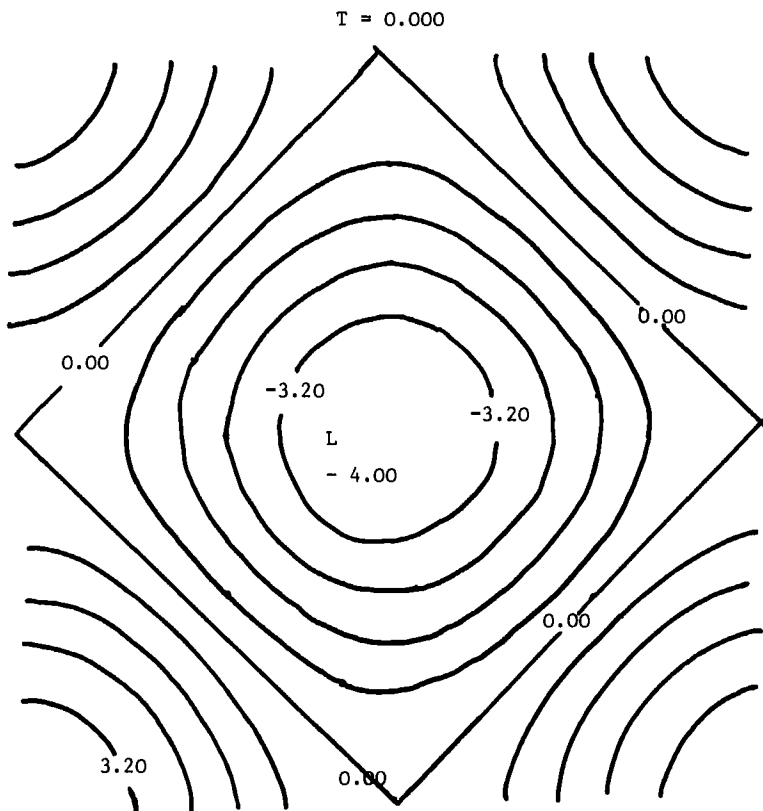


Figure 3

Orszag - Tang - Vortex
Magnetic Potential Contour for a ($b = \text{curl} a$)

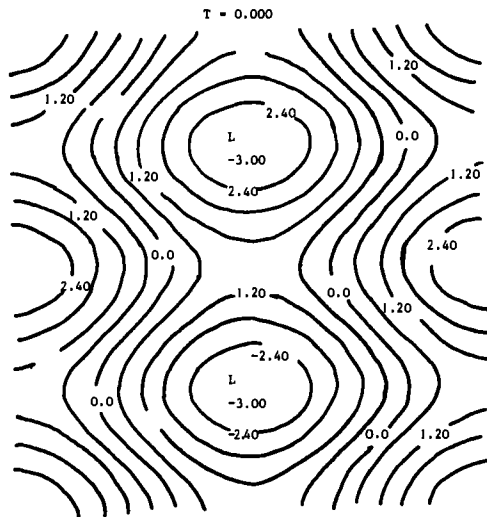


Figure 4

Kinetic Energy Spectrum at time $t = 4.000$ for the 2D. M.H.D. equation
with Orszag - Tang - Vortex as initial data.

Computation made by M. Meneguzzi on the C.R.A.Y. 1
of the N.C.A.R.

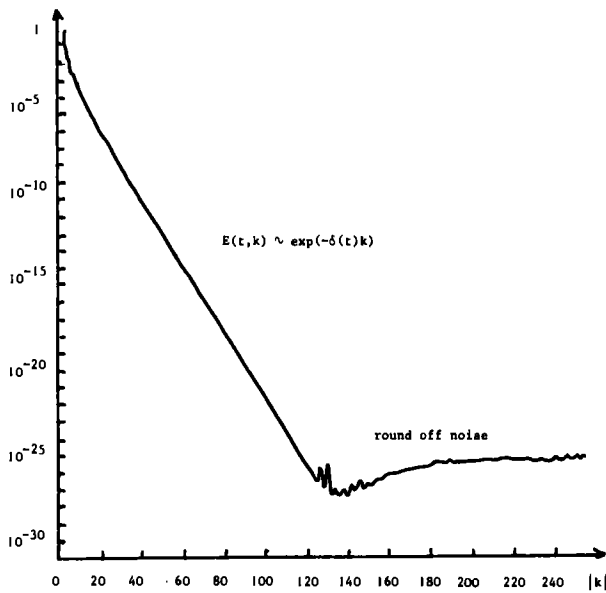


Figure 5

Kinetic Energy Spectrum at time $t = 6.000$ for the 2D. M.H.D. equation
with Orszag - Tang - Vortex as initial data. Computation made by
M. Meneguzzi on the C.R.A.Y. 1 of the N.C.A.R.

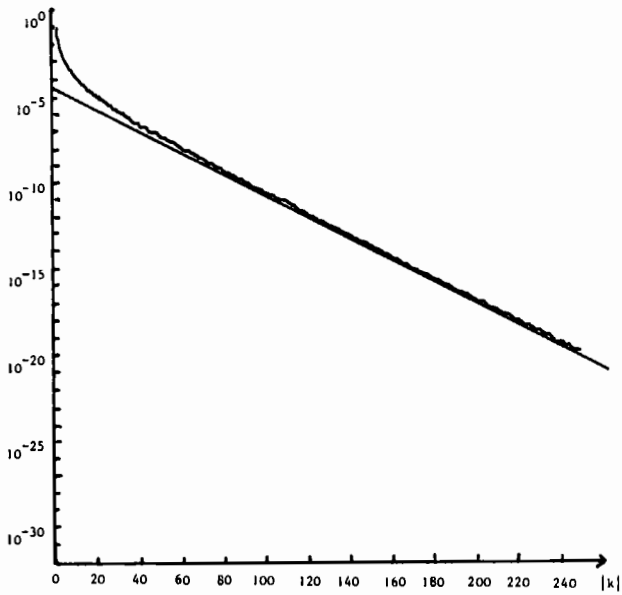
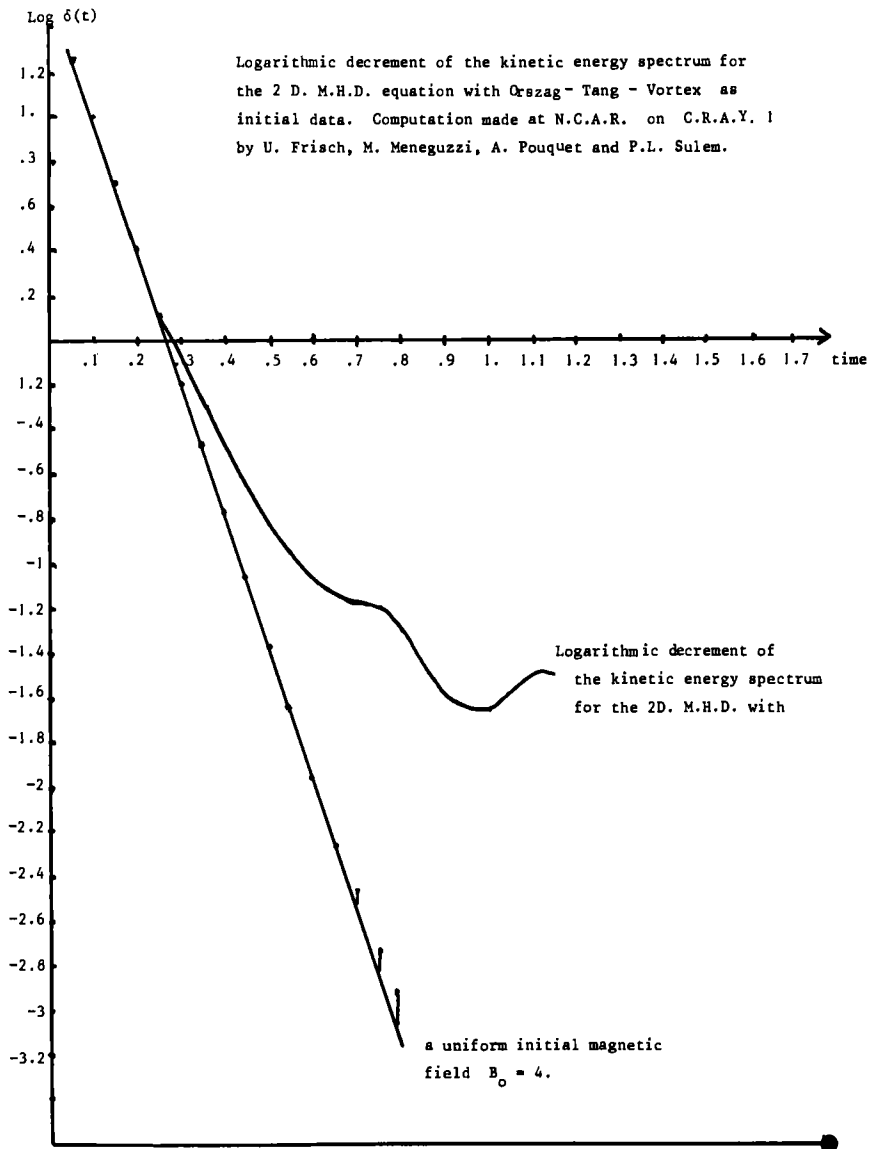
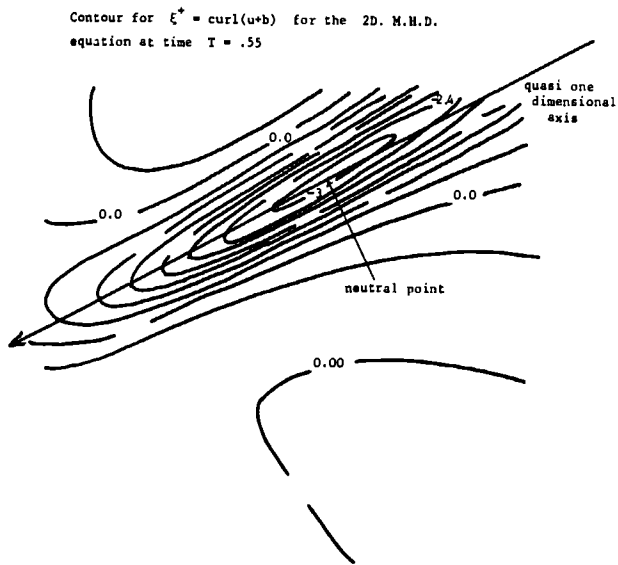
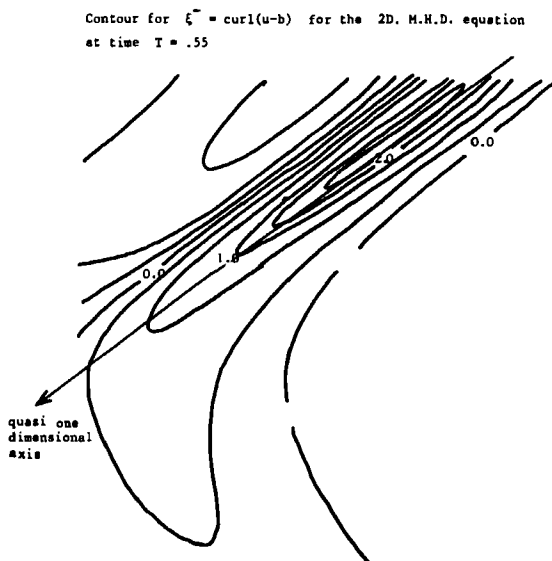


Figure 6

Figure 7

Figure 8Figure 9

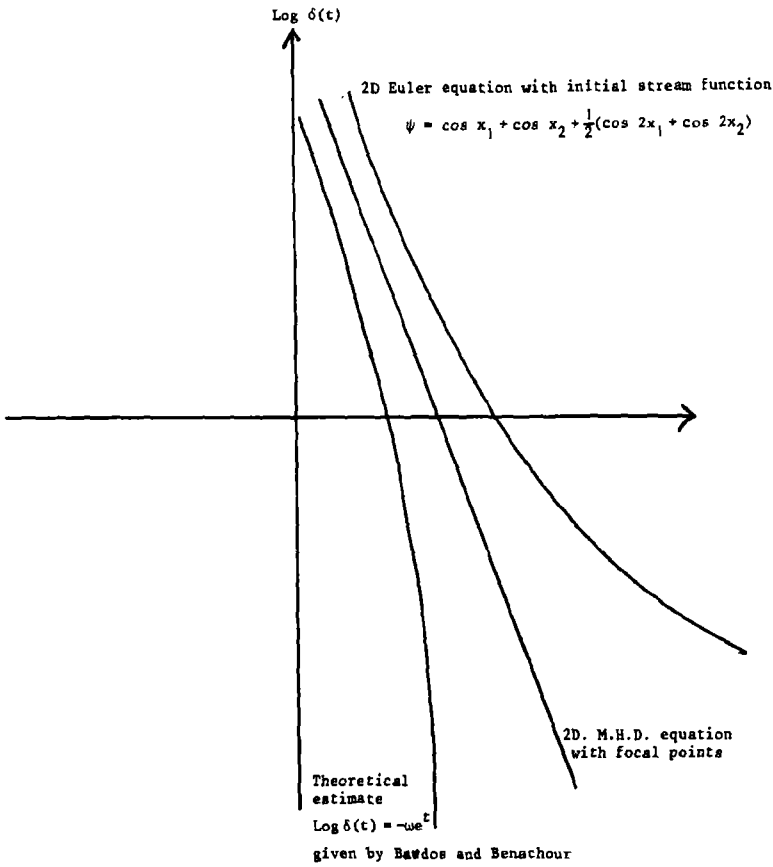


Figure 10

2. Euler equation in three space variables

The computation for the Euler equation in 3 space variables are made on a periodic solution with the following initial data

$$u_1(x_1, x_2, x_3) = u_2(-x_1, x_2, x_3) = \cos x_1 \sin x_2 \cos x_3 \quad u_3(x_1, x_2, x_3) = 0 \quad (52)$$

The quantity which is closely related to the appearance of Turbulence is the enstrophy $\Omega_1(t)$ given by

$$\Omega_1(t) = \int_X |\nabla \wedge u(x, t)|^2 dx = \sum_{k \in \mathbb{Z}^3} |k|^2 |\hat{u}(k, t)|^2 \quad (53)$$

The computation have been recently made in two different ways.

(1) For t real numerical computations are made using finite differences and the Fast Fourier Transform of Orszag.

(2) Keeping in mind that all the quantities

$$\Omega(t) = \sum_{k \in \mathbb{Z}^3} |k|^{2p} |\hat{u}(k, t)|^2 = \sum_{n=0}^{\infty} A_n^{(p)} t^{2n} \quad (54)$$

should blow up at the same type, one compute these expansions.

We will describe with some details this second method following closely the work of Morf, Orszag and Frisch [23]; one notices that the expansion (54) contains only even terms in t because the solution of the Euler Equation, with Taylor Green, Vortex for initial data is even in t . To obtain $A_n^{(p)}$ one needs only the derivatives of u at $t = 0$; one uses the formula

$$A_n^{(p)} = \frac{d}{dt^n} \left(\sum_{k \in \mathbb{Z}^3} |k|^{2p} (\hat{u}(k), \overline{\hat{u}(k)}) \right) \Big|_{t=0} \quad (55)$$

which involves the successive derivatives of $\hat{u}(k, t)$ for $t = 0$; these derivatives are given by the formula

$$\frac{\partial \hat{u}_\alpha}{\partial t} = - \sum_{\beta, \gamma=1}^3 k_\alpha (\delta_{\alpha, \beta} - \frac{k_\alpha k_\beta}{|k|^2}) \sum \hat{u}_\beta(p, 0) \cdot \hat{u}_\gamma(k-p, 0) \quad (56)$$

Due to the particular form of the initial data the authors [22] were able to compute the serie (54) up to the order 45. The coefficients $A_n^{(p)}$ are real and satisfy the relation : $A_n^{(p)} = (-1) A_{n+1}^{(p)}$. The numerical computation shows the existence of the limit

$$\lim_{n \rightarrow \infty} A_n^{(p)} / A_{n+1}^{(p)} = -5 \quad (57)$$

Which leads, for the complex extension of the Euler equation to a singularity at $t^2 = -5$. To locate the first real singularity one uses the Euler transform : $\omega = 6t^2/(t^2+5)$ the serie $\Omega_p(t) = \sum A_n^{(p)} t^{2n}$ is transformed in the serie $\tilde{S}_p(\omega) = \sum B_n^{(p)} \omega^n$. The coefficient of this serie are positive and the radius of convergence will determine the first real singularity. In the figure 11 we plot

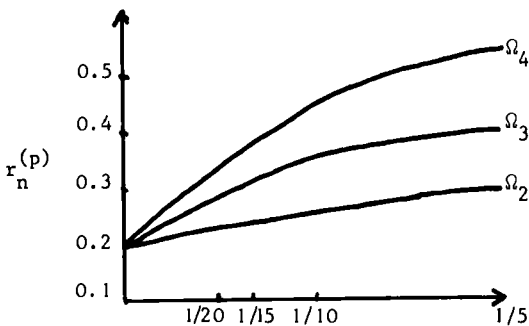


Figure 11 for [23]

the ratios

$$r_n^{(p)} = B_n^{(p)} / B_{n-1}^{(p)} \quad \text{for}$$

$1 < p < n$. For $p = 2, 3, 4$

these ratios decrease

monotonically with increasing

n and the extrapolation

to $n = \infty$ is consistent

with a common intersection

at $1/\omega(T^*) = r_\infty \approx 0.197$

or $t_* \approx 5.2$. An asymptotic

expansion of $r_n^{(p)}$ is then computed giving :

$$r_n^{(p)} \approx r_\infty (1 + (\gamma_p - 1)/n + \dots) \quad (58)$$

The formula (58) suggest for $\Omega_p(t)$ a power law behavior near T^* of the form $(T^* - t)^{-\gamma_p}$ with the critical exponents

$$\gamma_1 \approx 0.8, \gamma_2 \approx 4.2, \gamma_3 \approx 9.9, \gamma_4 \approx 16.$$

To locate the first real singularity one can also use Pade approximants instead of the Euler Transform, this is illustrated by the figure 12 where are plotted several curves concerning the computation of the enstrophy. The use of Euler

Transform or Padé approximants give also some informations concerning the complex singularities of the function $\Omega_1(t)$. These singularities are plotted on the figure 13.

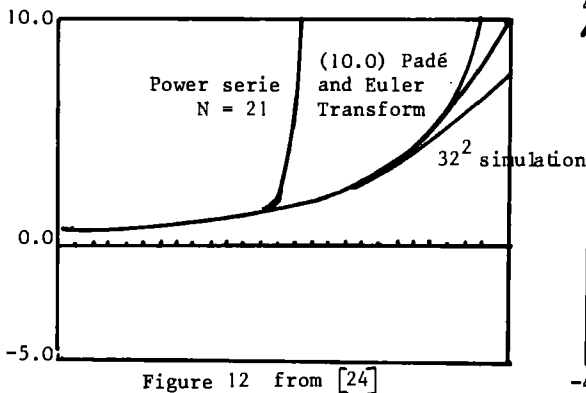


Figure 12 from [24]

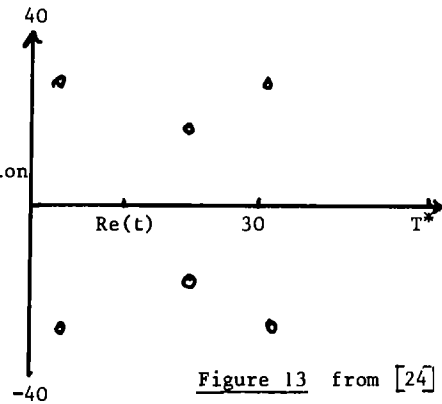


Figure 13 from [24]

Remark 2 : The existence of complex singularities approaching the real domain may be related (cf. Frisch and Morf [11] for a detailed discussion) to the appearance, before the turbulence, of intermittent burst. In fact a high pass filtering process, in the presence of complex singularity t_j will give a signal of the following form.

$$\Omega_R^1(t) \approx 1/\sqrt{R} \sum_j c_j e^{-R \operatorname{Im} t_j} \operatorname{Re} \left(\frac{\exp(-iR(t - \operatorname{Re} t_j) - i\pi/4)}{t - t_j} \right) \quad (59)$$

In (59) R denotes the size of the filter (the Fourier transform of $\Omega^1(t)$ has been made equal to zero for $|z| < R$), the asymptotic behaviour is given for $R \rightarrow \infty$ by a contour deformation around the half axis starting from the complex singularities of $\Omega^1(t)$.

To be really valid such a computation would require a detail knowledge of the behaviour of $\Omega^1(t)$ in the complex domain (This knowledge is not yet available) but in [11] one find some computations concerning other explicit examples like the Langevin equation or the Burger Hopf equation (with the use of the Hopf Cole transformation).

REFERENCES

- [1] Baouendi, S. and Goulaouic, C. : Remarks on the Abstract Form of Non linear Cauchy - Kovalevsky Theorems, Comm. Part. Diff. Eq. (1977) 1151-1162.
- [2] Bardos, C. : Existence et Unicité de la solution de l'équation d'Euler en deux dimensions, J. Math. Anal. and Appl. 40 (1972) 769-790.
- [3] Bardos, C. and Benachour, S. : Domaine d'Analyticité des solutions de l'équation d'Euler dans un ouvert de \mathbb{R}^n , Ann. Scuola Norm. Sup. di Pisa, Série IV Vol. IV, n° 4 (1977) 648-687.
- [4] Birkoff, G.: Helmholtz and Taylor Instability in Hydrodynamic Instability Proc. Symposium on Applied Math. 23, A.M.S.
- [5] Caffarelli, G. and Nirenberg, L. : To appear.
- [6] Chadam, J. : Asymptotics for $u = m^2 u + G(x, t, u, u_x, u_t)$, Global existence and decay, Annali. Scuola Norm. Sup. di Pisa, Vol. 26 (1972) 31-65.
- [7] Ebin, D. and Marsden, J. : Groups of Diffeomorphism and the motion of an incompressible fluid, Ann. of Math. 92 (1970) 102-163.
- [8] Foias, C. and Prodi, G. : Sur le comportement global des solutions non stationnaires des équations de Navier Stokes en dimension 2, Rend. Sem. Math. Padova 39 (1967) 1-34.
- [9] Foias, C. and Temam, R. : Some analytic and geometric Properties of the solution of the Navier Stokes Equations, Journal Math. Pure et Appl. Vol. 58 (1979) 339-368.
- [10] Foias, C., Frisch, U. and Temam, R. : Existence des solutions C^∞ des Equations d'Euler, C.R. Acad. Sc. Paris 280 A (1975) 505-508.
- [11] Frisch, U. and Morf, R. Intermittency in non linear dynamics and singularities at complex times, Phys. Rev. A.23.
- [12] Glimm, J. : Solutions in the large for non linear systems of equations, Comm. Pure Appl. Math. 18 (1965) 697-715.
- [13] Holder, E. : Über die Unberchränkte Fortsetzbarkeit einer Stetigen ebener Bewegung in einer unbegrenzten inkompressiblen Flüssigkeit Math. Z 37 (1933) 727-732.
- [14] Kato, T. and Fujita, M. : On the non stationary Navier Stokes Systems,

- Rend. Sem. Math. Univ. di Padova 32 (1962), 243-260.
- [15] Klainerman, S. : Global existence for non linear wave equation, Comm. Pure Appl. Math. 33 (1980), 43-101.
 - [16] Kolmogorov, A.N. : The local structure of turbulence in an incompressible viscous fluid for very large Reynolds number C.R. Ac. Sc. U.R.S.S. 30 : 301 also Soviet Physics Uspekhi 10 : 734.
 - [17] Kruckov, N.S. : First order quasilinear equations in several space independent variables, Math. Sbornik 81 (123) (1970), Math. U.R.S.S. Sbornik 10 (1970), 217-243.
 - [18] Ladyshenskaya, O.A. : Mathematical Theorie of incompressible viscous fluids Moscou 1961, Gordon - Breach New-York (1963).
 - [19] Leray, J. : Sur le mouvement d'un liquide visqueux emplissant l'espace, Acta Math. 63 (1934), 193-248.
 - [20] Lichtenstein, I. : Uber einige problem der Hydrodynamic, Math. Z.23 (1925).
 - [21] Lions, J.L. : Sur l'existence des solutions de Navier-Stokes C.R. Acad. Sc. Paris 248, (1959), 2847-2849.
 - [22] Meiron, D.I., Baker, G. and Orszag, S. : Analytic structure of Vortex Sheet Dynamics, Kelvin Helmholtz Instability, Preprint M.I.T. Department of Mathematics, Cambridge M.A. 012139.
 - [23] Morf, R., Orszag, S. and Frisch, U. : Spontaneous Singularity in Three-Dimensional, Inviscid, Incompressible Flow, Phys. Review Letters 44,9, (1980), 572-575.
 - [24] Morf, R., Orszag, S., Meiron, D., Frisch, U. and Meneguzzi, M. : Analytic structure of High Reynolds Number Flows, Seventh Conf. Numerical Methods in Fluid Dynamics Stanford June 1980, Springer Lecture Notes in Physics.
 - [25] Orszag, S. and Tang : Small Scale Structure of two dimensional Magneto-Hydrodynamic Turbulence, J. Fluid Mech. 90 (1979), 129- 136.
 - [26] Ovsjannikov, L. : Singular Operator in Banach Spaces, Dokl. Akad. Nank. U.R.S.S. 163 (1965), 819-822.
 - [27] Scheffer, G. : Navier Stokes Equation in a Bounded Domain, Comm. Math. Physic 73 (1980) 1-42.

- [28] Sulem, C. : Quelques resultats de regularité pour les equations de la Magneto-hydrodynamique, Comptes Rendus Acad. Sci. Paris 85 A (1977) 365-367.
- [29] Sulem, C., Fournier, J.P., Frisch, U. and Sulem,, P.L. : Remarques sur un modele unidimensionnel pour la Turbulence Magneto-Hydrodynamique, C.R. Acad. Sci. Paris t. 288 (12 mats 1979) Série A. 571-573.
- [30] Sulem, C., Sulem,, P.L., Bardos, C. and Frisch, U. : Finite Time Analyticity for the two and Three Dimensional Kelvin Helmholtz Instability, to appear in Comm. in Math. Physic.
- [31] Temam, R. and Strang, G. : Duality and relaxation in the variational problems of plasticity. Journal de Mécanique, Vol. 19, 3 (1980) 493-527.
- [32] Wolibner, W. : Un théorème sur l'existence du mouvement plan d'un fluide parfait et incompressible pendant un temps infiniment long, Math. Z. 37 (1933), 727-738.
- [33] Youdovitch, V.I. : The flow of a non stationary ideal and inviscid fluid Math. , Phys. and Numer Math. J. 6 (1965) 1032-1066.

This Page Intentionally Left Blank

FINITE PARAMETER APPROXIMATIVE STRUCTURE OF ACTUAL FLOWS

Ciprian Foias and Roger Temam

Analyse Numérique et Fonctionnelle
CNRS et Université Paris-Sud
Bâtiment 425
91405 - Orsay Cedex (France)

INTRODUCTION

Most approaches to the mathematical understanding of the onset of turbulence (e.g. [6], [4], [10], [1]) assume that the complexity of the dynamical system $\{S(t)\}_{t \geq 0}$ associated to the initial value problem for the Navier-Stokes equations in some adequate functional space H (formed by divergence free velocities fields) is increasing in an almost universal monotonic way, with sudden jumps at some threshold values of a physically significant parameter. However this universal pattern (although displayed by adequately constructed models) was not established for actual flows, for which even the definition of $\{S(t)\}_{t \geq 0}$ is not yet satisfactory. The only known fact, which is relevant to this problem, is that for plane flows (with stationary boundary conditions and driving forces) all its trajectories $S(t)u_0$ converge, for $t \rightarrow \infty$, to a set $X_{\max} \subset H$ of finite Hausdorff dimension d_{\max} , for which a rather explicit, upper estimate $d(1/\nu)$, as function of the kinematic viscosity ν , exists ([9], [3]§.6). Although $d(1/\nu) \nearrow \infty$ for $1/\nu \nearrow \infty$ it is not yet known whether $d_{\max} \rightarrow \infty$ for $1/\nu \rightarrow \infty$. On the other hand the change from a laminar to a turbulent flow of a fluid in motion is often so fast that it is hard to conceive that its mathematical understanding lies only on the asymptotic behaviour of $u(t)$ for $t \rightarrow \infty$ and thus on the structure of a (still hypothetical, if the flow is not plane) attracting set $X \subset H$.

A way by which nature could satisfy all the guessed approaches is given in the following

Conjecture. Any trajectory $u(t)$ is "very rapidly" converging (in H) to a manifold M of finite dimension d such that $d \rightarrow \infty$ for $1/\nu \rightarrow \infty$.

To prove that this conjecture is "generically" true seems to be outside the grasp of the present mathematical techniques. Strangely enough, we can give a rigorous positive answer to an approximative form of the conjecture. In order to state this answer let us recall that H (see the next §.1) is a real Hilbert space and that $1/2$ of the square $|u|^2$ of the norm of $u \in H$ represents the kinetical energy of the flow $u \in H$. (Actually we assume that the density of the fluid is $\rho = 1$). This is our

Answer (see §.3 below). For given $0 < E_1 \ll 1 \ll E_0$ and $0 < \tau < 1$ there exists a manifold M of dimension $d < \infty$ such that for all trajectories $u(t)$ with initial data $|u(0)|^2 \leq E_0$, the set of the t 's for which

- (1) distance (in H) from $u(t)$ to $M \leq E_1^{1/2}$ is of measure $\geq 1 - \tau$ on any interval $\subset (0, \infty)$ of length $= 1$.

Actually this answer is not unexpected since on any set $V_R \subset H$ of velocities fields such that

$$\int_{\Omega} |\operatorname{curl} u|^2 dx \leq R$$

the map $S(t_0)$ (namely the displacement $u(0) \mapsto u(t_0)$ along the trajectory $u(t)$) is a well defined one to one compact diffeomorphism ([2], Ch. III). However we shall show that

$$(2) \quad d \sim (\log 1/E_1)^n$$

(for E_0 , τ and ν fixed), which depends on a more particular structure of $S(t_0)$, where $n = 2$ for plane flows and $n = 3$ otherwise.

The quantity E_1 might represent the unavoidable experimental error in the estimation of the energy of a flow or even might represent a much more basic entity comparable with the mean kinetical energy of a molecule of the given fluid at rest. In this case it is clear that differentiating two flows u, v such that $|u-v|^2 \leq E_1$, although mathematically meaningful is void of any physical basis. But because of (2) even this microscopical E_1 will yield a reasonable non astronomical function of E_0 , τ and ν .

The construction of M , given in §.3, also provides some supplementary information on its position in H allowing us to show (in §.4) that the behaviour of d , as function of $1/\nu$, is consistent with the conjecture.

§.1. PRELIMINARIES ⁽¹⁾

We shall consider the Navier-Stokes equations

$$(1.1) \quad u_t + (u \cdot \nabla)u = \nu \Delta u + \nabla p + f, \quad \nabla \cdot u = 0 \quad \text{in } \Omega \times (0, \infty)$$

where Ω is an open connected bounded set $\subset \mathbb{R}^n$ ($n=2$ or 3) such that its boundary Γ is a compact manifold of class C^2 of dimension $n-1$, and Ω is locally located on one side of $\partial\Omega$. In order to simplify the exposition we shall first consider only the boundary problem

$$(1.2) \quad u|_{\Gamma} \equiv 0.$$

As usual, let H and V (see [3].§.2 for more details) denote the closure in $L^2(\Omega)^n$ and $H^1(\Omega)^n$, respectively, of

$$\mathcal{V} = \{v : v \in C_0^\infty(\Omega)^n : \nabla \cdot v = 0\},$$

⁽¹⁾ See [3].§§.1-3 for more details.

and

$$(1.3) \quad \begin{cases} Au = -P\Delta u & \text{for } u \in \mathcal{D}(A) \stackrel{\text{def}}{=} V \cap H^2(\Omega)^3 \\ B(u,v) = P[(u \cdot \nabla)v] & \text{for } u, v \in \mathcal{D}(A) \end{cases}$$

where $H^\ell(\Omega)$ ($\ell=1,2$) denote the Sobolev L^2 -space of order ℓ and P denotes the orthogonal projection of $L^2(\Omega)^n$ onto H . There exists an orthonormal basis $\{w_m\}_{m=1}^\infty$ of H such that $Aw_m = \lambda_m w_m$ ($m=1,2,\dots$), $0 < \lambda_1 \leq \lambda_2 \leq \dots$ and

$$(1.4) \quad \lambda_m \sim c_1 m^{2/n} \quad (2)$$

(where c_1 as the others c_j 's in the sequel ($j=2,3,\dots$) denote positive constants depending only on Ω). The orthogonal projection in H onto $\mathbb{R}w_1 + \dots + \mathbb{R}w_m$ will be denoted by P_m ($m=1,2,\dots$; $P_0 = 0$). We set

$$|u|^2 = \int_{\Omega} \sum_{j=1}^n u_j^2 dx, \quad \|v\|^2 = \int_{\Omega} \sum_{k=1}^n \left(\frac{\partial v_j}{\partial x_k} \right)^2 dx \quad (u \in H, v \in V);$$

$|u|$ will be the norm on H , $\|v\|$ that on V and (u_1, u_2) and $((v_1, v_2))$ will denote the corresponding scalar products. The operator B is continuous from $\mathcal{D}(A) \times V$ and $V \times \mathcal{D}(A)$ to H as well as from $V \times V$ into the dual V' of V . For many more fine estimates on B we refer to [3], §.1. Also B enjoys the following basic orthogonality property

$$(1.5) \quad B(u,v), v = 0$$

whenever the left hand side makes sense.

With the above notations the initial value problem for (1.1), (1.2) can be written in H , namely

$$(1.6) \quad \frac{du}{dt} + v \cdot Au + B(u,u) = Pf \quad \text{for } t > 0, \quad u(0) = u_0$$

where $u_0 \in H$ is given. For the only sake of simplifying the presentation in the sequel we shall assume that $Pf = 0$, i.e. that the forces are potential.

The following are well known classical results (see, for instance, [5], [6] or [11]):

The problem (1.6) has a (weak) solution $u(\cdot)$ such that

$$(1.7) \quad u(\cdot) \in C(0, \infty; H_{\text{weak}}) \cap L_{\text{loc}}^2(0, \infty; V).$$

(2) Actually we shall use only the fact that $\lambda_m \geq c_2 m^{2/n}$ for all $m = 1, 2, \dots$ where c_2 is suitably chosen, $0 < c_2 < c_1$.

Actually, because of our assumption on f we have

$$(1.8) \quad |u(t)| \leq |u_0| \quad (t \geq 0) \quad \text{and} \quad \int_0^\infty \|u(t)\|^2 dt \leq \frac{1}{2\nu} |u_0|^2$$

If $u_0 \in V$ then there exists an interval $[0, t(u_0))$, where

$$(1.9) \quad t(u_0) \geq t_0(\|u_0\|) \stackrel{\text{def}}{=} c_3 \nu^3 \|u_0\|^{-4}$$

on which all (weak) solution coincide with a (strong or regular) solution, i.e.

$$(1.10) \quad u(\cdot) \in C(0, t(u_0); V_{\text{strong}}) \cap L^2_{\text{loc}}(0, \infty; \mathfrak{D}(A)) .$$

For such a strong solution the definition

$$(1.11) \quad S(t)u_0 = u(t) \quad 0 \leq t < t(u_0)$$

makes sense. Moreover in case $n=2$, or in case $n=3$ but if $\|u_0\|^4 \leq c_4 \nu^4$, one has $t(u_0) = \infty$ and any weak solution is unique and obviously is regular on any interval $[t_0, \infty)$ ($t_0 > 0$).

Among the other regularity properties of $S(t)u_0$ let us mention only the fact that $S(t)u_0$ is a $\mathfrak{D}(A)$ -valued analytic function of $t \in (0, t(u_0))$ such that

$$(1.12) \quad |AS(t)u_0| \leq \begin{cases} c_5 \frac{\|u_0\|}{\sqrt{\nu t}} \\ c_6 \frac{|u_0|}{\sqrt{\nu t}} \end{cases} \quad \text{for } 0 < t \leq t_0(\|u_0\|)$$

(see [3], §.3).

§.2. THE SQUEEZING PROPERTY

In the sequel, a basic role will be played by the following squeezing property of $S(t)$, which we obtained in [3], §.5 :

For any $r > 0$, $t \in [\frac{1}{2} t_0(r), t_0(r)]$ and $u, v \in V$, $\|u\|, \|v\| \leq r$, we have either

$$(2.1) \quad |(I - P_m)(S(t)u - S(t)v)| \leq |P_m(S(t)u - S(t)v)|$$

or

$$(2.2) \quad |S(t)u - S(t)v| \leq c_1(r, \nu) e^{-c_2(r, \nu) \lambda_{m+1}^{1/2}} |u - v| ,$$

whenever

$$(2.3) \quad \lambda_{m+1} \geq c_3(r, v),$$

where $0 < c_1, c_2, c_3 < \infty$ do not depend on t, u, v or m .

A careful perusal of the proof (given in [3], §.5) of this property yields the estimates

$$(2.4) \quad \begin{cases} c_1(r, v) \leq c_7 \\ c_2(r, v) \geq c_8(r/v)^{-5/2} \\ c_3(r, v) \leq c_9(r/v)^5 \end{cases}$$

where the constants c_8 and c_9 are not dimensionless. The estimates (2.4) are valid for both cases $n=2$ and $n=3$. (Actually for $n=2$, the exponents $5/2$ and 5 in (2.4) can be decreased to 2 and 4 , respectively).

In case $n=3$ it can be shown that $c_8 \geq c_{10} \lambda_1^{1/8}$ and $c_9 \leq c_{10} \lambda_1^{-1/4}$ with some dimensionless constants c_{10}, c_{11} . Introducing the variables

$$(2.5) \quad \begin{cases} K = \lambda_{m+1}^{1/2}, \quad K_1 = \lambda_1^{1/2} \\ \epsilon = \nu r^2 |\Omega|^{-1} \quad \text{and} \quad K_d = (\epsilon/\nu^3)^{1/4} \end{cases}$$

the relation (2.3) can be written under the form

$$(2.6) \quad \frac{K}{K_1} \geq \left(\frac{K_d}{K_1} \right)^5 c_{12} \quad (\text{with } c_{12} \text{ dimensionless}),$$

which can be interpreted as showing that the frequencies K for which the squeezing property of $S(t)$ holds are lying far inside the dissipative spectrum of the flow. This strongly suggests that the exponents in (2.4) must be much nearer to $1/2$ and 1 , respectively.

§.3. THE APPROXIMATIVE STATIONARY MANIFOLD

Let $0 < E_1 \ll 1 \ll E_0$, $0 < \tau \ll 1$, be such that

$$(3.1) \quad E_0 \geq (2c_3)^{1/2} \tau^{1/2} \nu^{5/2}, \quad \frac{E_0}{\nu \tau \lambda_1 E_1} \geq c_1^{(3)}$$

⁽³⁾ $c_1^{(3)} = \frac{1}{2c_7^2} \exp(2c_8 c_9^{1/2})$. Actually we can drop these conditions, but this would make the expression (3.2) much more complicated.

Then we have the following result, loosely stated in the Introduction :

Theorem 1 - There exist m and a map $\psi : P_m H \mapsto (I - P_m)H$ satisfying the properties

$$(3.2) \quad m \leq c_{13} (E_0 / \sqrt{\tau})^{\frac{n(n+2)}{4}} (\log \frac{E_0}{\sqrt{\tau} \lambda_1 E_1})^n$$

$$(3.3) \quad |\psi(P_m u) - \psi(P_m v)| \leq |P_m u - P_m v| \quad (u, v \in H),$$

and such that for any weak solution $u(\cdot)$ of (1.6) the set

$$(3.4) \quad \{t \in (0, \infty) : \text{distance (in } H) \text{ from } u(t) \text{ to } M_\psi \leq E_1^{1/2}\},$$

(where $M_\psi \subset H$ is the graph of ψ) is of measure $\geq 1 - \tau$ on any interval $\subset (0, \infty)$ of length $= 1$, and $n (=2 \text{ or } 3)$ is the dimension of Ω .

Before passing to the proof, let us notice that M_ψ is obviously a Lipschitz manifold of dimension m , which can be called an approximative stationary manifold of (1.6).

Proof of Theorem 1.

Let $u(\cdot)$ be a (weak) solution of (1.6) such that $|u_0|^2 \leq E_0$. Then

$$\forall \int_t^{t+1} \|u(s)\|^2 ds \leq \frac{1}{2} |u(t)|^2 \leq \frac{1}{2} |u_0|^2 \leq \frac{1}{2} E_0.$$

Thus for any $r > 0$,

$$r^2 |\{s \in (t, t+1) : \|u(s)\| > r\}| \leq \int_t^{t+1} \|u(\tau)\|^2 d\tau \leq \frac{1}{2\sqrt{\tau}} E_0$$

and consequently if

$$(3.5) \quad r \geq (E_0 / \sqrt{\tau})^{1/2}$$

then

$$(3.6) \quad |\{s \in (t, t+1) : \|u(s)\| \leq r\}| \geq 1 - \tau/2 \quad \text{for any } t \geq 0.$$

Let $t_0 = t_0(r)$ and choose a subset $M(m)$ of

$$S(t_0) \{u_0 \in V : \|u_0\| \leq r\}$$

maximal under the property

$$(3.7) \quad |(I - P_m)(u - v)| \leq |P_m(u - v)| \quad (u, v \in M(m))$$

where m satisfies

$$(3.8) \quad \lambda_{m+1} \geq c_9(\gamma/\nu)^{n+2}.$$

Then, by virtue of the squeezing property (see §.2), we have that the distance (in H) of $u(s+t_0) = S(t_0)u(s)$ to $M(m)$ is bounded by

$$(3.9) \quad \mu_m = c_7 \frac{2\gamma}{\lambda_1^{1/2}} \exp\left(-c_8 \nu^{\frac{n+2}{2}} \gamma^{\frac{n+2}{2}} \lambda_{m+1}^{1/2}\right)$$

whenever $\|u(s)\| \leq r$. But for any $t \geq t_0$, we have obviously (by (3.6))

$$|\{s+t_0 \in (t, t+1) : \|u(s)\| \leq r\}| \geq 1 - \tau/2$$

so that if $t_0 < 1 - \frac{\tau}{2}$, then for all $t \geq 0$:

$$|\{s \in (t, t+1) : \|u(s)\| \leq r\}| \geq 1 - \tau/2 - t_0.$$

If we assume now that

$$(3.10) \quad \mu_m \leq E_1^{1/2} \quad \text{and} \quad t_0(=t_0(r)) \leq \tau/2$$

then it follows that the set of t 's such that the distance of $u(t)$ to $M(m)$ is $\leq E_1^{1/2}$ has measure $\geq 1 - \tau$ on any interval $(t, t+1) \subset (0, \infty)$. By (3.7), $M(m)$ is the graph of the function $\psi_0 : P_m M(m) \mapsto (I - P_m)H$ defined by

$$\psi_0(P_m u) = (I - P_m)u \quad (P_m u \in P_m M(m)),$$

which moreover satisfies the condition

$$(3.11) \quad |\psi_0(P_m u) - \psi_0(P_m v)| \leq |P_m u - P_m v| \quad (P_m u, P_m v \in P_m M(m)).$$

By virtue of the Kirsbaum extension theorem (see [8]) ψ_0 can be extended on the whole $P_m H$ to a function ψ satisfying (3.3). Thus it remains only to verify (3.2). To this aim we first note that due to the relation (1.9) and the first relation (3.1), the second condition (3.10) is a direct consequence of (3.5). Similarly, due to the second condition (3.1), (3.5), and the first condition (3.10) imply (3.8). Finally we must have to impose (3.5) and the first condition (3.10). We define r by replacing (3.5) by an equality. Replacing the first relation (3.10) by an equality and taking (1.4) into account, we get (3.2).

We remark that in case $n=2$ the approximative stationary manifold M_ψ can be chosen such that any solution eventually becomes and remains within distance $E_1^{1/2}$ of M_ψ .

§.4. THE BEHAVIOUR FOR $\nu \rightarrow 0$.

We want to study the behaviour of the dimension m of approximative stationary manifolds for $\nu \rightarrow 0$. In order to avoid very difficult boundary layer problems we make the remark that all the previous considerations remain valid if Ω is replaced by $Q = [0, L]^n$ where $L > 0$ is fixed, and the boundary problem (1.2) is replaced by the periodic conditions

$$(4.1) \quad u(0, \dots) = u(L, \dots), \dots, u(\dots, 0) = u(\dots, L).$$

Since the Navier-Stokes equations are Galilean invariant, we can assume, by a change of the reference frame, that the solutions verify also

$$(4.2) \quad \int_Q u \, dx = 0.$$

Therefore H and V become the closure in $L^2(Q)^n$ and $H^1(Q)^n$, respectively, of the space of all \mathbb{R}^n -valued trigonometric polynomials $w(x)$ such that

$$\int_Q w \, dx = 0 \quad \text{and} \quad \nabla \cdot w \equiv 0.$$

All the other definitions and properties remain essentially unchanged. Actually in this case much more concrete algebraic structure is present. For instance the λ_m are now of the form $4\pi^2 L^{-2} |k|^2$ with $k \in \mathbb{Z}^n$, $k \neq 0$ and $\nu Au + B(u, u)$ has the following property.

For any $m = 1, 2, \dots$ there exists an isometric operator $U : H \rightarrow (I - P_m)H$ such that

$$(4.3) \quad U[\nu Au + B(u, u)] = \lambda [\nu A Uu + B(Uu, Uu)] \quad (u \in D(A))$$

where λ is some adequate constant such that

$$(4.4) \quad 0 < \lambda \leq c_{14} \lambda_m^2.$$

One defines $U w_j$ ($j = 1, 2, \dots$) according to the following scheme : First choose the first $\ell (= 1, 2, \dots)$ such that $\pi 4^{\ell+1} L^{-2} > \lambda_m$, then if $\lambda_j = 4\pi L^{-2} |k|^2$ chose j such that $\lambda_j = 4\pi L^{-2} |Q^\ell k|^2$ and set $U w_j = w_j$ (Here some supplementary care is necessary in order to avoid the complication due to the fact that the multiplicity of λ_j or λ_j is not simple. Since we intend to study this self similarity phenomena elsewhere, we don't insist with the details of the proof.)

Finally, for $u_0 \in V$ the solutions of (1.6) satisfy, in case $n=2$, besides the relations (1.8), the relations

$$(4.5) \quad \|u(t)\| \leq \|u_0\| \quad (t \geq 0) \quad \text{and} \quad \int_0^\infty |Au(t)|^2 dt \leq \frac{1}{2\nu} \|u_0\|^2.$$

Theorem 2 - Let $n=2$ and let M_ν be an approximative stationary manifold of (1.6). If E_1/E_0 and τ are sufficiently small (for instance $\tau < 1/3$ and $E_1/E_0 < 1/36$, then

$$(4.6) \quad m_\nu = \text{dimension of } M_\nu \rightarrow \infty \text{ for } \nu \rightarrow 0.$$

Proof :

Let $M_\nu = M_{\psi_\nu}$ where $\psi_\nu = \psi$ enjoys the properties given in Theorem 1.

If $u, v \in H$ and their distance to M_{ψ_ν} is $\leq E_1^{1/2}$, then there exist u_*, v_* in M_{ψ_ν} with

$$|u - u_*| \leq E_1^{1/2}, \quad |v - v_*| \leq E_1^{1/2}$$

so that

$$|u - v| \leq 2E_1^{1/2} + |P_{m_\nu}(u_* - v_*)| + |(I - P_{m_\nu})(u_* - v_*)|$$

$$|u - v| \leq 2E_1^{1/2} + 2|P_{m_\nu}(u_* - v_*)|$$

$$|u - v| \leq 2E_1^{1/2} + 2|P_{m_\nu}(u - u_*)| + 2|P_{m_\nu}(v - v_*)| + 2|P_{m_\nu}(u - v)|$$

$$(4.7) \quad |u - v| \leq 6E_1^{1/2} + 2|P_{m_\nu}(u - v)|.$$

Therefore if $u(\cdot)$ and $v(\cdot)$ are solutions of (1.6) with $u(0) = u_0$, $v(0) = v_0$ satisfying $|u_0|, |v_0| \leq E_0$, it follows from (4.7) and (3.4) (for $u(\cdot)$ and $v(\cdot)$) that

$$|\{t \in (0,1) : |u(t) - v(t)| \leq 6E_1^{1/2} + 2|P_{m_\nu}(u(t) - v(t))|\}| > 1 - 2\tau$$

Since for $v(\cdot)$ we can take $v(t) \equiv 0$, we obtain that for any solution $u(\cdot)$ of (1.6) such that $|u_0| \leq E_0$ the following holds

$$(4.8) \quad |\{t \in (0,1) : |u(t)| \leq 6E_1^{1/2} + 2|P_m u(t)|\}| > 1 - \tau.$$

Now we assume that there exists a sequence $\nu_j \searrow 0$ such that $m_{\nu_j} \leq m < \infty$ for all $j=1,2,\dots$. Let now ν be any of these ν_j and let U be the isometric operator considered in (4.3) and let $u_0 = UE_0^{1/2} w_1$. We denote by $\tilde{u}(\cdot)$ the

solution of (1.6) with initial data $E_0^{1/2} w_1$. Then by (4.3)

$$\frac{dU\tilde{u}}{dt} + \lambda [vAU\tilde{u} + B(U\tilde{u}, U\tilde{u})] = 0$$

and therefore $u(t) = (U\tilde{u})(t/\lambda)$ is a solution of (1.6) with initial data $u_0 (= E_0^{1/2} U w_1)$, $|u_0| = E_0^{1/2}$. Thus (4.8) is valid for this $u(\cdot)$. But because U^H is orthogonal on P_m^H it follows that $P_{m_v} u(t) \equiv 0$ and therefore (4.8) becomes

$$(4.9) \quad |\{t \in (0,1) : |u(t)| \leq 6E_1^{1/2}\}| > 1 - \tau.$$

But, by (1.8) and (4.5),

$$\begin{aligned} |u(t)|^2 - |u_0|^2 &= -2v \int_0^t \|u(\tau)\|^2 d\tau \geq -2vt \|u_0\|^2 \\ &= -2vt \|E_0^{1/2} U w_1\|^2 = -2vt E_0 \|U w_1\|^2 \\ &= -2vt E_0 4\ell \geq -2vt E_0 L^2 \lambda_m / \pi \end{aligned}$$

so that (4.9) implies :

$$(4.10) \quad |\{t \in (0,1) : E_0 \leq 36E_1 + 2vt E_0 \frac{L^2 \lambda_m}{\pi}\}| \geq 1 - \tau.$$

From (4.10) it follows that there exists a $t \in (0, 2\tau)$ such that

$$E_0 \leq 36E_1 + 2vt E_0 \frac{L^2 \lambda_m}{\pi}$$

and consequently

$$E_0 \leq 36E_1 + 4v\tau E_0 \frac{L^2 \lambda_m}{\pi} \quad \text{for all } v = v_j \quad (j=1,2,\dots).$$

We conclude with the contradiction $E_0 \leq 36E_1$. This proves the Theorem.

Actually the Theorem is also valid if $n=3$ but the proof must be modified in order to work with small solutions in V , and consequently the condition $E_0/E_1 < 1/36$ has to be replaced by $E_0/E_1 \leq c_{15}$ with an adequate constant c_{15} depending on the constant c_4 (see §.1).

REFERENCES

- [1] M.J. Feigenbaum, J. Stat. Phys. 21, 669 (1979); Phys. Lett. 74A, 375 (1979); see also these proceedings.

- [2] C. Foias, Solutions statistiques des équations de Navier-Stokes
Cours au Collège de France (1974).
- [3] C. Foias and R. Temam, Some analytic and geometric properties of the
solutions of the evolution Navier-Stokes equations
J. Math.pure et appl., 58 (1978) pp.339-369.
- [4] E. Hopf, A mathematical example displaying feature of turbulence
Comm.pure and appl. Math., 1 (1948), pp.303-322.
- [5] O.A. Ladyzenskaya, The mathematical theory of viscous incompressible flows
Gordon-Breach, New York (1969).
- [6] L.D. Landau and E.M. Lifshitz, Fluid Mechanics
Pergamon, Oxford (1959).
- [7] J.L. Lions, Quelques méthodes de résolution des problèmes aux limites non
linéaires
Dunod-Gauthier-Villars, Paris (1969)
- [8] J.H. Wells and L.R. Williams, Imbeddings and extensions in Analysis
Springer-Verlag, Heidelberg-New York.
- [9] J. Mallet-Parret, Negatively invariant sets of compact maps and an extension
of a theorem of Cartwright
J. Diff. Equations, 22 (1976), pp. 331-348.
- [10] D. Ruelle and F. Takens, On the nature of turbulence
Comm. Math. Phys., 20 (1971), pp.167-192 ;
Comm. Math. Phys., 23 (1971), pp.343-344.
- [11] R. Temam, Navier-Stokes equations. Theory and Numerical Analysis
North-Holland, Amsterdam, New York (1977).

This Page Intentionally Left Blank

THE ROLE OF CHARACTERISTIC BOUNDARIES IN THE ASYMPTOTIC SOLUTION OF THE NAVIER-STOKES EQUATIONS

F. A. Howes

Department of Mathematics
University of California, Davis
Davis, California 95616
U.S.A.

We discuss several boundary value problems for the steady Navier-Stokes equations in two dimensions when the appropriate Reynolds number Re is very large. The essential idea is that certain curves in the flow field on which the conditions of no-slip and impermeability are imposed are characteristic curves (in the mathematical sense) for the system of differential equations governing the conservation of momentum. This observation, coupled with some of the author's recent results on singularly perturbed elliptic boundary value problems, lead to several mathematically rigorous estimates on the behavior of solutions of the Navier-Stokes equations as $Re \rightarrow \infty$. For example, our approach shows why the thickness of the boundary layer in Prandtl's theory is of order $Re^{-1/2}$, without having to resort to the usual heuristic arguments.

1. STATEMENT OF THE PROBLEM

Let us consider the two-dimensional, steady flow of a homogeneous, viscous, incompressible fluid near a plane boundary. The equations which govern the velocity and pressure fields, written in dimensional form, are the Navier-Stokes equations (conservation of momentum)

$$\begin{aligned}v \Delta^* u^* &= u^* u_{\xi}^* + v^* u_{\eta}^* + \rho^{-1} p_{\xi}^* \\v \Delta^* v^* &= u^* v_{\xi}^* + v^* v_{\eta}^* + \rho^{-1} p_{\eta}^*,\end{aligned}$$

and the continuity equation (conservation of mass)

$$u_{\xi}^* + v_{\eta}^* = 0;$$

cf. [1; Chap. 3]. Here $\xi(u^*)$ and $\eta(v^*)$ are the tangential and normal components of direction (velocity), respectively, $\Delta^* = \partial^2/\partial \xi^2 + \partial^2/\partial \eta^2$, ρ is the constant density, p^* is the dynamic pressure (that is, the difference of the actual pressure from the hydrostatic pressure) and $\nu = \mu/\rho$ is the kinematic viscosity.

In order to study this system more effectively, we render the variables dimensionless by referring distances to a representative length L and velocities to a uniform speed U , that is,

$$\begin{aligned}x &= \xi/L, \quad y = \eta/L \\u &= u^*/U, \quad v = v^*/U, \quad p = p^*/(\rho U^2).\end{aligned}$$

These changes of variable allow us to rewrite the equations of motion as

$$\begin{aligned}(N-S) \quad & \epsilon \Delta u = uu_x + vu_y + p_x \\& \epsilon \Delta v = uv_x + vv_y + p_y \\(C) \quad & u_x + v_y = 0,\end{aligned}$$

where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ and $\epsilon = 1/\text{Re} = \nu/(UL)$, for Re the dimensionless group known as the Reynolds number. The problem is completed by requiring that u and v satisfy the conditions of no-slip ($u=0$) and impermeability ($v=0$) at a solid boundary in the flow field, and that the velocity field (u, v) approach the uniform flow away from such boundaries.

Our study of this boundary value problem will concern the case of high Reynolds number flow $\text{Re} \rightarrow \infty$ (or equivalently, $\epsilon \rightarrow 0$), and so we turn now to several approximate forms of the system which utilize the smallness of ϵ .

2. EULER FLOWS

If we formally set $\epsilon = 0$ in (N-S) we obtain the Euler equations of motion for an inviscid (ideal) fluid, namely

$$(E) \quad \bar{u} \bar{u}_x + \bar{v} \bar{u}_y + \bar{p}_x = 0$$

$$\bar{u} \bar{v}_x + \bar{v} \bar{v}_y + \bar{p}_y = 0$$

$$(C) \quad \bar{u}_x + \bar{v}_y = 0;$$

cf. [1; Chap. 6]. The system (E) is first-order, whereas the corresponding system (N-S) is second-order, and consequently, the system (E), (C) together with the original boundary conditions is an over-determined problem. In order to obtain a meaningful solution of (E), (C), one traditionally drops the condition of no-slip at a solid boundary Σ on the grounds that the tangential component \bar{u} of an ideal flow can slide, that is, $\bar{u} \neq 0$ on Σ . Clearly, then, the solution $(\bar{u}, \bar{v}, \bar{p})$ of the Euler problem cannot be a uniformly valid approximation to the solution of the viscous ($\epsilon > 0$) problem, since in the neighborhood of a solid boundary u and \bar{u} can differ by an amount of order one. A striking manifestation of this failure of the Euler flow to model the actual flow is given by d'Alembert's paradox; cf. [1; Chap. 5].

3. PRANDTL'S BOUNDARY LAYER THEORY

In 1904 L. Prandtl [11] (see also [12; Chap. 7]) was able to resolve d'Alembert's paradox, and at the same time, to derive a simplified form of the Navier-Stokes equations whose solutions modelled quite closely the behavior of the actual flow near a solid boundary. The starting point of his investigation was the no-slip boundary condition, which led him to reason that there is a thin layer near the boundary in which viscous forces outweigh inertial and pressure forces in the direction normal to the flow. That is, near a solid boundary Σ the gradients u_x , v_x , and v_y change very little relative to the normal gradient u_y . As a result, some terms in (N-S) could be neglected, at least near Σ . This thin layer in which Prandtl's simplified equations are valid was termed the "boundary layer".

For example, we can derive Prandtl's boundary layer equations in the case of a boundary parallel to the x -direction and located at $y = 0$ if we proceed formally as follows. We begin by rescaling the y and v variables by setting

$$\sigma = y/\epsilon^\alpha, \quad \hat{v} = v/\epsilon^\beta, \quad \hat{u} = u, \quad \theta = x, \quad \hat{p} = p,$$

for positive constants α and β to be determined. Then we substitute into the continuity equation (C) and obtain

$$0 = u_x + v_y \equiv \hat{u}_\theta + \epsilon^{\beta-\alpha} \hat{v}_\sigma.$$

Physical arguments suggest that $\alpha = \beta$ is the only acceptable choice of α and β . Proceeding next to the momentum equations (N-S) we obtain

$$(N-S') \quad \epsilon \hat{u}_{\theta\theta} + \epsilon^{1-2\alpha} \hat{u}_{\sigma\sigma} = \hat{u} \hat{u}_\theta + \hat{v} \hat{v}_\sigma + \hat{p}_\theta$$

$$\epsilon \hat{v}_{\theta\theta} + \hat{v}_{\sigma\sigma} = \hat{u} \hat{v}_\theta + \hat{v} \hat{v}_\sigma + \epsilon^{-1} \hat{p}_\sigma.$$

Once again, physical arguments related to the matching of the flow near the boundary with the Euler flow suggest that the constant α must be $1/2$. The second equation in (N-S') implies that to lowest order in ϵ , $\hat{p}_0 = 0$ in the boundary layer, that is, $\hat{p} \equiv \bar{p}$, the inviscid pressure. If we now let $\epsilon \rightarrow 0$ in (N-S') with $\alpha = 1/2$ and rewrite the limiting equations in the original, dimensionless variables, then we obtain Prandtl's celebrated boundary layer equation

$$(P) \quad \epsilon u_{yy} = uu_x + vu_y - \bar{u}\bar{u}_x.$$

This equation is parabolic with y a space-like and x a time-like variable. The choice of $\alpha = 1/2$ implies the formal result that the thickness of the boundary layer is proportional to $\epsilon^{1/2}$, that is, the size of the region of nonuniformity of the Euler approximation is of order $\epsilon^{1/2}$.

4. MATHEMATICAL BOUNDARY LAYER THEORY

Several mathematicians have been able to prove, under various assumptions on the flow field, that the solution of the Navier-Stokes equations (N-S) reduces to the solution of the Prandtl equation (P) as $\epsilon \rightarrow 0$ near a solid boundary, and that the solution of (N-S) reduces to the "reasonable" solution of the Euler system (E) as $\epsilon \rightarrow 0$ away from such curves. We mention only the work of Oleinik [10], Nickel [9] and Fife [4] which the reader can consult for details and further references.

5. SOME SINGULAR PERTURBATION THEORY

In order to motivate our discussion of the Navier-Stokes equations, which to a certain extent obviates the use of Prandtl's approximation, let us consider several simple singularly perturbed problems.

The first example is the problem

$$(I) \quad \begin{aligned} \epsilon \Delta u &= mu, & (x,y) \text{ in } D(0;1), \\ u &= \varphi(x,y), & (x,y) \text{ on } S^1, \end{aligned}$$

where $D(0;1)$ is the open unit disk in \mathbb{R}^2 , S^1 is the unit circle, and m is a positive constant. This problem can be solved by separation of variables, and we find that its unique solution $u = u(x,y;\epsilon)$, to transcendently small terms, can be written as

$$u(x,y;\epsilon) = O(|\varphi(x,y)| \exp[(x^2+y^2-1)(m\epsilon^{-1})^{1/2}])$$

for (x,y) in $\overline{D(0;1)}$. In other words,

$$\lim_{\epsilon \rightarrow 0} u(x,y;\epsilon) = 0$$

in every compact subset of $D(0;1)$, and u differs from 0 by an amount of order one in a region (boundary layer) of order $\epsilon^{1/2}$ since the boundary data φ is not necessarily zero. We note that $\bar{u} \equiv 0$ is the solution of the reduced (Euler) equation obtained from (I) by formally setting $\epsilon = 0$, and that the order of the differential equation drops from two to zero.

The next example shows that the boundary layer can also be of order ϵ , namely

$$(II) \quad \begin{aligned} \epsilon \Delta u &= xu_x + yu_y + u, & (x,y) \text{ in } D(0;1), \\ u &= \varphi(x,y), & (x,y) \text{ on } S^1. \end{aligned}$$

If we set $\epsilon = 0$, we obtain

$$xu_x + yu_y = -\bar{u},$$

which is Euler's relation for homogeneous functions of degree (-1) (cf. [2; Chap. 1]). Its only singularity-free solution is $\bar{u} \equiv 0$. From the exact solution

or from the theory of the author [5] we see that the solution of (II) can be expressed as

$$u(x,y;\epsilon) = O(|\varphi(x,y)| \exp[(x^2+y^2-1)(1-\delta)/\epsilon])$$

for (x,y) in $\overline{D(0;1)}$ where $0 < \delta < 1$. Here we observe that in setting $\epsilon = 0$ the order of the differential equation drops from two to one. The resulting first-order equation (*) has characteristic (base) curves given by the solutions of

$$\frac{dx}{ds} = x, \quad \frac{dy}{ds} = y,$$

that is, they are the one-parameter family of straight lines $y = cx$. It is clear that they exit $D(0;1)$ along S^1 nontangentially, a fact which implies that the boundary layer is of order ϵ (cf. [8] or [5,6]) since the gradient terms in (II) actually determine the boundary layer structure.

We consider next a problem in which the right-hand side contains gradient terms and yet, the solution behaves near the boundary as if these terms were absent. It is (cf. [5])

$$(III) \quad \begin{aligned} \epsilon \Delta u &= y u_x - x u_y + u, \quad (x,y) \text{ in } D(0;1), \\ u &= \varphi(x,y), \quad (x,y) \text{ on } S^1. \end{aligned}$$

The characteristic curves of the reduced ($\epsilon = 0$) equation are given by the solutions of the system

$$\frac{dx}{ds} = y, \quad \frac{dy}{ds} = -x,$$

that is, they are the family of concentric circles $x^2 + y^2 = c^2$. Clearly the boundary S^1 is itself a characteristic curve, and from the asymptotic form of the solution in $\overline{D(0;1)}$

$$u(x,y;\epsilon) = O(|\varphi(x,y)| \exp[(x^2+y^2-1)/\epsilon^{1/2}]),$$

we see that the boundary layer is of order $\epsilon^{1/2}$. In other words, near the boundary the solution u of (III) behaves like the solution of problem (I) (with $m=1$). The gradient terms in (III) have no effect since the boundary is characteristic, a fact that can be expressed analytically by the relation

$$(y, -x) \cdot \nabla(x^2+y^2-1) = 0.$$

This situation is typical in cases where the boundary of the region is a characteristic surface of the first-order reduced equation; cf. [7] and [5,6].

The final example of this section has a solution which displays behavior analogous to that encountered in Examples (II) and (III). The problem is

$$(IV) \quad \begin{aligned} \epsilon \Delta u &= -u_x \equiv (-1, 0) \cdot \nabla u, \quad (x,y) \text{ in } R, \\ u &= \varphi(x,y), \quad (x,y) \text{ on } \partial R, \end{aligned}$$

where $R = \mathcal{J} \times \mathcal{J}$, $\mathcal{J} = (-1, 1)$ and $\partial R = \Sigma_+ \cup \Sigma_- \cup \Xi_+ \cup \Xi_-$, for $\Sigma_{\pm} = \{\pm 1\} \times \mathcal{J}$ and $\Xi_{\pm} = \mathcal{J} \times \{\pm 1\}$. The boundary of the square R can also be represented as the set $\{(x,y): F(x,y) = 0\}$ for $F(x,y) = -(1-x^2)(1-y^2)$. Clearly we have that

$$\text{along } \Sigma_-: (-1, 0) \cdot \nabla F > 0$$

and

$$\text{along } \Sigma_+: (-1, 0) \cdot \nabla F < 0.$$

However, along the sides Ξ_+ and Ξ_- we have that

$$(-1, 0) \cdot \nabla F \equiv 0.$$

Thus, in view of the two previous examples, we might expect that the solution of Example (IV) exhibits boundary layer behavior along the sides Σ_- , Ξ_+ and Ξ_- , and that the thickness of the boundary layer region is of order $\epsilon^{1/2}$ along Σ_+ (Ξ_+ and Ξ_-). This turns out to be the correct asymptotic behavior of the solution $u = u(x,y;\epsilon)$; cf. [3], [6]. We note finally that along the side Σ_+ where $(-1, 0) \cdot \nabla F < 0$ there is no boundary layer, and that the boundary data along Σ_+

actually determine the limiting solution in R . Indeed, it follows from the theory that

$$\lim_{\epsilon \rightarrow 0} u(x, y; \epsilon) = \varphi(1, y)$$

in each compact subset of R .

The examples of this section are meant to illustrate the idea that it is the interaction of the characteristic curves of the reduced ($\epsilon = 0$) differential equation with the boundary of the region that determines the location and the size of the boundary layers. With this as background, we return finally to the Navier-Stokes system.

6. CHARACTERISTIC CURVES ASSOCIATED WITH THE NAVIER-STOKES EQUATIONS

The reduced system corresponding to the Navier-Stokes equations is the Euler system

$$(E) \quad \begin{aligned} uu_x + vu_y &= -p_x \\ uv_x + vv_y &= -p_y. \end{aligned}$$

(We have dropped the overbars for simplicity.) This is a quasilinear system of first-order equations whose principal part is the velocity field (u, v) (cf. [2; App. 1 of Chap. 2]); consequently, the characteristic equations associated with (E) are

$$\frac{dx}{ds} = u, \quad \frac{dy}{ds} = v, \quad \frac{du}{ds} = -p_x, \quad \frac{dv}{ds} = -p_y.$$

Thus, for small values of ϵ we expect, based on the considerations of §5, that the boundary layer phenomena in the Navier-Stokes system are related to the interaction of these characteristic curves with the geometry of a solid boundary in the flow field. In other words, if the boundary of a plane region is described by the equation $F(x, y) = 0$, then the sign of the inner product

$$(*) \quad (u(x, y), v(x, y)) \cdot \nabla F(x, y)$$

should determine the location and the size of the boundary layers.

For example, the conditions of no-slip ($u = 0$) and impermeability ($v = 0$) make the inner product $(*)$ vanish irrespective of the behavior of ∇F , and so we know that such a boundary is a characteristic curve for the Euler system (E). We expect then that the size of the boundary layer along such a curve is proportional to $\epsilon^{1/2}$; cf. Examples (III) and (IV).

Similarly, for parallel flow from left to right in the rectangular region $D = (0, L) \times (0, L) \subset \mathbb{R}^2$, the entrance (exit) section $x = 0$ ($x = L$) is noncharacteristic for the u -equation, and since the flow enters D along $x = 0$ we should prescribe data for u there. However, since the flow exits along $x = L$ we expect a boundary layer there of order ϵ for the u -variable (cf. Example (IV)). As regards the walls $y = 0, L$, we note that they are characteristic for the u -variable, and so we expect boundary layers there of order $\epsilon^{1/2}$ for the u -variable.

The situation is reversed for the v -variable. The sections $x = 0, L$ are now characteristic for it, leading us to expect boundary layers of order $\epsilon^{1/2}$ there; while the walls $y = 0, L$ are noncharacteristic. Depending on boundary conditions such as suction or blowing, we expect to find boundary layers of order ϵ along $y = 0, L$ for the v -variable.

These considerations can be extended to more complicated geometries in \mathbb{R}^2 and to problems in three dimensions. The idea is essentially the same: it is the interaction of the characteristics of the associated Euler system with the solid boundaries in the flow field that determines the location and the size of the boundary layers in the actual flow.

ACKNOWLEDGMENTS. The author wishes to thank the National Science Foundation for its financial support and Ida Mae Orahod for typing the manuscript.

REFERENCES

- [1] Batchelor, G. K., *An Introduction to Fluid Dynamics* (Cambridge U. Press, Cambridge, 1970).
- [2] Courant, R. and Hilbert, D., *Methods of Mathematical Physics*, vol. II (Interscience, New York, 1962).
- [3] Eckhaus, W. and de Jager, E. M., Asymptotic Solutions of Singular Perturbation Problems for Linear Differential Equations of Elliptic Type, *Arch. Rational Mech. Anal.* 23(1966), 26-86.
- [4] Fife, P. C., Toward the Validity of Prandtl's Approximation in a Boundary Layer, *ibid.* 18(1965), 1-13.
- [5] Howes, F. A., Singularly Perturbed Semilinear Elliptic Boundary Value Problems, *Comm. in P.D.E.* 4(1979), 1-39.
- [6] Howes, F. A., Some Singularly Perturbed Nonlinear Boundary Value Problems of Elliptic Type, *Proc. Conf. Nonlinear P.D.E.'s in Appl. Sci. and Engrg.*, ed. by R. L. Sternberg, pp. 151-166, Marcel Dekker, New York, 1980.
- [7] Kamenomostvskaya, S. L., On Equations of Elliptic and Parabolic Type with a Small Parameter Multiplying the Highest Derivatives (in Russian), *Mat. Sbornik* 31(1952), 703-708.
- [8] Levinson, N., The First Boundary Value Problem for $\epsilon \Delta u + A(x,y)u_x + B(x,y)u_y + C(x,y)u = D(x,y)$ for Small ϵ , *Ann. Math.* 51(1950), 428-445.
- [9] Nickel, K., Prandtl's Boundary Layer Theory from the Viewpoint of a Mathematician, *Ann. Rev. Fluid Mech.* 5(1973), 405-428.
- [10] Oleinik, O., On a System of Equations in Boundary Layer Theory, *U.S.S.R. Comp. Math. Math. Phys.* 3(1963), 650-673.
- [11] Prandtl, L., Über Flüssigkeitsbewegungen bei sehr kleiner Reibung, *Proc. Third Int. Math. Congress Heidelberg 1904*, pp. 484-494, Teubner, Leipzig, 1905. Translated as NACA TM 452.
- [12] Schlichting, H., *Boundary Layer Theory*, 2nd English ed. (McGraw-Hill, New York, 1960).

SOME APPROACHES TO THE TURBULENCE PROBLEM

J. Mathieu and D. Jeandel
Laboratoire de Mécanique des Fluides (L.A. 263)

C. M. Brauner
Laboratoire de Mathématiques-Informatique-Systèmes
Ecole Centrale de Lyon
69130 Ecully
FRANCE

This paper devoted to non-linear topics deals with a very specific case encountered in fluid mechanics. The problem is examined from two points of view. At first, the turbulent field is considered, this gives rise to discussions on the crucial problem of closure which is examined in several ways. Connections between turbulence problems and mechanical problems treated in a statistical way are recalled. Even where a second order closure method is concerned, mathematical developments have to be introduced to solve a set of equations supplemented by boundary conditions. In the second part of this paper mathematical developments are presented, they correspond to fairly simple cases of closure, notwithstanding they bring to light the mathematical difficulties to treat these problems in an accurate way.

PRELIMINARY REMARKS

This lecture is intended to discuss several aspects of a non-linear problem which is encountered in fluid dynamics and called "turbulence". The basic problem is conveniently settled; for an incompressible fluid the functions \vec{U} and P of \vec{X} and t are linked to each other through the Navier-Stokes and continuity equations. The existence and uniqueness of a solution for a set of boundary and initial conditions is conjectured, not demonstrated, except for two-dimensional space. It can also be postulated that the solutions exhibit a highly oscillating character which would lead to a statistical approach to this problem. Even in this form the problem is not trivial because we have to find, through a non-linear equation, statistical information at any point within a domain, starting from initial and boundary conditions which are also given statistically. Numerical approaches require very large computers with adequate nets; the relative influence of viscous and inertial terms strongly depends on the distance to a wall which tends to introduce several scalings. On the other hand, the numerical scheme should not alter the properties of the basic equations whose solutions can be strongly oscillating. To clarify

this problem mathematical research has been initiated in several directions. L. D. Landau and Hopf have postulated that the basic system could generate sets of stable solutions from bifurcation points. So far, this theory cannot globally be supported even though situations presenting some analogies have been encountered. In the case of rotating cylinders (Taylor's experiments), W. I. Yudowich and V. Velte have proved the existence of bifurcations which lead to new solutions which are not as symmetrical as the initial one. D. Ruelle and F. Takens conjecture that the transition could be explained by the existence of "strange attractors" for unsteady solutions of the Navier-Stokes equations.

In fact, no accurate information is available where three dimensional space is concerned. We can also consider that when turbulence mechanisms appear, the physical meaning of deterministic solutions is lost, giving rise to stable statistical distributions. So far, we ignore the exact meaning of these statistical solutions in three dimensional space. The equations for the joint characteristic function are known but mathematical difficulties previously encountered reappear in fairly similar forms.

However, some recent findings should open the way to a more grounded theory of homogeneous turbulence.

I. AN OVERALL VIEW OF TURBULENCE PROBLEMS

I.1. Introduction

Most physical phenomena are represented by mathematical functions (depending on space-time variables) which generally satisfy non-linear equations, however in most cases a linear treatment can be considered as fairly convenient. For example, classical electromagnetic phenomena are controlled by linear equations. On the contrary, regarding fluid dynamic equations, the role played by the non-linear terms appears determinant. In some previous works very slow motions have been considered and the role played by the convective terms neglected, but so drastic an assumption concerned a very closed field of applications. However, recent utilizations of the Stokes operator to solve the Navier-Stokes equations are remarkable.¹ Fluid dynamic problems giving rise to crucial non-linear mechanisms must be considered as most instructive in a conference typically devoted to non-linear topics.

In fluid dynamics, turbulence is one of the most challenging effects in connection with non-linear mechanisms. In the subsequent chapters, we accept the fluid motion to be conveniently represented by the Navier-Stokes equation;

in an Eulerian frame we have:

$$\frac{\partial U_i}{\partial t} + U_j \frac{\partial U_i}{\partial x_j} = - \frac{1}{\rho} \frac{\partial p}{\partial x_i} . \quad (1.1)$$

The fluid is supposed to be newtonian; if so, the non-linearity is only a consequence of convective terms. Besides, it is conjectured that a large gap exists between the smallest turbulent scales and the molecular characteristic sizes so that the above equation should be reliable for any turbulent structures (even for the smallest dissipative structures).

During the last decade, a great effort has been made with sophisticated methods in order to understand turbulence mechanisms; this trend should not dissimulate historical development more specially devoted to find tractable ways for an unclosed problem.²⁻⁴ Starting from a deterministic equation for the instantaneous motion a statistical formalism has been introduced with the moments about the mean value of the velocity. These moments have been treated in a more or less sophisticated manner. For many years, such a theory prevailed; so far, it is still very useful for technical applications. On the other hand, the increasing capability of large computers has led scientists to perform direct numerical simulations of turbulent fields using Navier-Stokes equations; computations start from both adequate initial and boundary conditions. The complete properties of the basic equations are obviously considered with a numerical scheme which should not alter them. Even though the exact influence of such alterations is always open to debate, available results are interesting. Anyway, the turbulence problem can be considered in several different ways.

We can first examine what may be origin of the random nature of many phenomena and, for example, shortly recall the physical equations which govern the associated density probability functions in several fields. Laying aside the Schrodinger equation, which is introduced in quantum mechanics from a deterministic approach, but whose solutions are interpreted in a probabilistic way, two main patterns of random phenomena may be considered. The first group is controlled by deterministic equations; however, random initial conditions generate random solutions (Liouville equations belong to this group). In the second group, initial conditions are formulated in a deterministic way but the random process is introduced through random coefficients in the basic equations (for example the Langevin equation for Brownian motion).

We must immediately remark that turbulence exhibits a rather typical case of non-linear equations. It can be conjectured that initial and boundary conditions determine the flow field at any time; so far, the uniqueness of the solutions has not yet been ensured except for very restricted conditions which are not encountered in the problem under consideration which is typically located in three dimensional space. From this point of view the role played by the bifurcation theory proposed by L. D. Landau and E. Hopf could become illuminating but at the present time this approach, which is in progress, is not sufficiently worked out to permit a straightforward analysis of the Navier-Stokes equations. In the case of unsteady solutions the role of "strange attractors" has been examined by D. Ruelle and F. Takens, but only partial conclusions have been given which cannot be extended to three dimensional space.

For the first group, with Hamiltonian systems, the Liouville equation is well adapted to predict the states of the system; a probability density function is introduced in phase space:

$$f[p_i, q_i, t] \quad (1.2)$$

and it is found that:

$$\frac{\partial f}{\partial t} + \dot{p}_i \frac{\partial f}{\partial p_i} + \dot{q}_i \frac{\partial f}{\partial q_i} = 0 \quad (1.3)$$

with

$$\dot{p}_i = \frac{\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i} \quad (1.4)$$

H stands for the Hamiltonian of the system.

For the second group, we must recall that Brownian motion leads to a Fokker-Planck equation provided that the memory of the system has faded enough.

When we have to deal with a reactive medium, the concentration equation can be written as:

$$\frac{\partial c_i}{\partial t} + v_j \frac{\partial c_i}{\partial x_j} = a_{c_i} \frac{\partial^2 c_i}{\partial x_j^2} + \gamma[c_1 \dots c_1 \dots] \quad (1.5)$$

Through the production term $\gamma[C_1 \dots C_i]$ a non-linear mechanism generally occurs. Some connection can also be found between the Fokker-Planck equation and this one. In fact, U_j plays the role of a random coefficient in the concentration equations but the Markovian character of the system is invalidated.

The problem under consideration can also be examined from the first point of view. We must treat a problem with deterministic equations and random initial conditions. The equations relative to kinematic properties of the fluid, and those concerning the concentration of the chemical species have to be considered as a blocked system having no random coefficient,

$$\begin{aligned} \frac{\partial C_i}{\partial t} + U_j \frac{\partial C_i}{\partial x_j} &= \alpha_{C_i} \frac{\partial^2 C_i}{\partial x_j^2} + \gamma[\dots] \\ \frac{\partial U_i}{\partial t} + U_j \frac{\partial U_i}{\partial x_j} &= -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \gamma \frac{\partial^2 U_i}{\partial x_j^2} \\ \frac{\partial U_j}{\partial x_j} &= 0 \end{aligned} \quad (1.6)$$

Moreover, the existence and the uniqueness of a solution is supposed for sufficiently smooth initial conditions, but so far, this assumption is not supported by mathematical considerations. However, this approach has first been developed by Hopf (1952),⁵ Monin (1969),⁶ Lundgreen (1967),⁷ and more recently by Dopazo and O'Brien (1975).⁸ After cumbersome algebraic computations a hierarchy of equations is found. At N given points in phase space they give the probability density functions to find fixed values of velocities and concentrations.

In order to simplify this section we propose classifying the existing methods for the modelling of turbulence into four groups.⁹ Developments which are reported herein principally concern homogeneous fields, with constant mean velocity gradients (or even null: isotropic turbulence) with adequate initial conditions. From these theories, inhomogeneous effects can only be considered as small; moreover the Reynolds number is supposed large, so that each group of turbulent structures located in a range of the spectrum should play a typical role. Besides, for any statistical properties of the field, the departure from a normal law is rather moderate. In other words, the methods

will be available for patterns of flows which can be relaxed to homogeneity and gaussianity. Characteristic length and time scales are introduced; they are necessary for the precision of the two concepts of homogeneity and stationarity.

I.2. Physical Space Methods

I.2.1. General Concepts and Second Order Closures

Disregarding what occurs in the spectral space where the role of turbulent structures is identified as a function of their sizes and directions, the instantaneous velocity \vec{U} is split into two parts

$$\vec{U} = \vec{\bar{U}} + \vec{u} \quad .$$

At first, using the Navier-Stokes equation it is found that:

$$\frac{\partial \bar{U}_i}{\partial t} + \bar{U}_j \frac{\partial \bar{U}_i}{\partial x_j} = - \frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_i} + \frac{1}{\rho} \frac{\partial}{\partial x_j} \left[\frac{\partial \bar{U}_i}{\partial x_j} - \overline{u_i u_j} \right] \quad . \quad (1.7)$$

At this stage, a first order closure can be introduced. For instance we may write:

$$\overline{u_i u_j} = v_T \left[\frac{\partial \bar{U}_i}{\partial x_j} + \frac{\partial \bar{U}_j}{\partial x_i} \right] - \frac{\delta_{ij}}{3} \overline{u_l u_l} \quad (1.8)$$

where v_T is a predetermined function of \vec{x} and t .

The introduction of moments about mean values does not presume anything about the origin of random mechanism which is considered by its consequences only. Where second order closures are concerned the previous equation is supplemented by a rate equation for the second order moment only:

$$\frac{\partial \overline{u_i u_j}}{\partial t} - \frac{\rho}{\rho} \left[\frac{\partial \bar{U}_i}{\partial x_j} + \frac{\partial \bar{U}_j}{\partial x_i} \right] + \overline{u_i u_k} \frac{\partial \bar{U}_j}{\partial x_k} + \overline{u_j u_k} \frac{\partial \bar{U}_i}{\partial x_k} - 2v \frac{\partial \bar{U}_i}{\partial x_k} \cdot \frac{\partial \bar{U}_j}{\partial x_k} = 0 \quad . \quad (1.9)$$

The whole turbulent flow is characterized by adequate length and time scales. In fact, a spectral equilibrium is assumed so that turbulent scales are linked to each other: for example, the macro and microscales are related through the turbulent Reynolds number. Moreover for large Reynolds numbers small

dissipative structures are supposed isotropic so that:

$$\overline{v \frac{\partial u_i}{\partial x_k} \cdot \frac{\partial u_j}{\partial x_k}} = \frac{2}{3} \bar{\varepsilon} \delta_{ij} \quad (1.10)$$

Finally we have:

$$\frac{\overline{\partial u_i u_j}}{\partial t} - \frac{p}{\rho} \left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right] + \overline{u_i u_k} \frac{\partial \bar{u}_j}{\partial x_k} + \overline{u_j u_k} \frac{\partial \bar{u}_i}{\partial x_k} - \frac{2}{3} \bar{\varepsilon} \delta_{ij} = 0 \quad (1.11)$$

The open nature of the problem is displayed by both the mean dissipation rate and the rate of strain pressure correlations. The equation which controls the dissipation rate can be written:

$$\frac{\partial \bar{\varepsilon}}{\partial t} = -2v \overline{\left[\frac{\partial u_i}{\partial x_k} \cdot \frac{\partial u_j}{\partial x_k} \cdot \frac{\partial u_i}{\partial x_j} \right]} - 2v^2 \overline{\left[\frac{\partial^2 u_i}{\partial x_j \partial x_k} \cdot \frac{\partial^2 u_i}{\partial x_j \partial x_k} \right]} \quad (1.12)$$

where for the sake of simplification, dimensionless forms are introduced

$$P_{ij} = \frac{1}{\varepsilon} \frac{p}{\rho} \left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right]$$

$$\Psi = \frac{q^2}{\varepsilon} \left[2v \overline{\frac{\partial u_i}{\partial x_k} \cdot \frac{\partial u_j}{\partial x_k} \cdot \frac{\partial u_i}{\partial x_j}} + 2v^2 \overline{\frac{\partial^2 u_i}{\partial x_i \partial x_k} \cdot \frac{\partial^2 u_i}{\partial x_j \partial x_k}} \right] \quad (1.13)$$

At this stage, no additional information is requested from the Navier-Stokes equation, and the closure is obtained by another means. The method which has been extensively developed by Lumley¹⁰ in its more perfect form considers the whole turbulent structure as a complex material, the properties of which have to be determined by experience. Therefore P_{ij} and Ψ_{ij} are supposed to be dependent on conveniently selected arguments. The deviation tensor b_{ij} being introduced as

$$b_{ij} = \frac{\overline{u_i u_j}(t)}{q^2(t)} - \frac{\delta_{ij}}{3} \quad (1.14)$$

it can be written that

$$\Psi = \Psi [b_{ij}, \overline{q^2}, \bar{\varepsilon}, \frac{\partial \bar{u}_i}{\partial x_k}] \quad (1.15)$$

"A priori" the introduction of functional forms underlines a non-local dependence, the memory of the structures being correlated with their sizes. For dissipative structures it is fairly convenient to admit that their characteristic time is small with respect to a characteristic time defined from the mean flow, for example, $\lambda/\tau_L = \partial \bar{u}_x / \partial x_m$. In fact, Lumley proposes that the two previous functionals Ψ and \bar{b}_{ij} are considered as simple functions, which assume that the memory of turbulent structures has faded enough. It is a crude assumption which can only be supported by its consequences.

Besides, if we admit that the dissipative structures are not directly strained by the mean field (the amount energy flux is transformed without alteration through the inertial range, but during these exchanges it can be conjectured that turbulence structures lose most of their orientation), the anisotropy is considered as globally taken into account through b_{ij} , (disregarding the mean velocity gradient which generates such an anisotropy). We can then write:

$$\Psi = \Psi [b_{ij}, \overline{q^2}, \bar{\varepsilon}] \quad (1.16)$$

With a suitable choice of arguments, we can admit Ψ to be an isotropic function of b_{ij} , $\overline{q^2}$ and $\bar{\varepsilon}$. Pipkin and Rivlin have demonstrated that such a function can only depend on the invariants (with respect to the orthogonal group) which are determined from b_{ij} , $\overline{q^2}$ and $\bar{\varepsilon}$.

For a small degree of anisotropy the previous function can be expanded with respect to the anisotropic parameter:

$$\Psi = f[\bar{\varepsilon}, \overline{q^2}] + g[\bar{\varepsilon}, \overline{q^2}] II + \dots \quad (1.17)$$

More precisely we find that:

$$\Psi \simeq A + B II \Rightarrow \frac{\partial \Psi}{\partial \bar{\varepsilon}} = - [A + B II] \frac{\bar{\varepsilon}^{-2}}{\overline{q^2}} \quad (1.18)$$

where II is the second invariant of b_{ij} , the constants must be adjusted from experimental data.

The second function P_{ij} can be determined in a similar but more complex way, the pressure term having linear and non-linear relations to the fluctuating field.

$$-\frac{\Delta p}{\rho} = 2 \frac{\partial \bar{u}_i}{\partial x_j} \cdot \frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \cdot \frac{\partial u_j}{\partial x_i} \quad (1.19)$$

These methods have extensively been developed during the last decade in several forms by Launder, Jeandel,

I.2.2. Computing Methods for Complex Flows

We will take advantage of this section devoted to modelling methods in physical space to indicate some progress in computing approaches for complex turbulent flows.

The patterns of bounded flows under consideration are strongly non homogeneous in more or less extended areas; the rate equation for the mean flow has to be coupled with a more or less complex turbulence model. Where a first gradient approach is used with V_T as a given function of space and time variables, we must deal with both the momentum equation for the mean field and the continuity equation.

The plane channel flow simplifications result in the dimensionless basic equation:

$$R_o \frac{\partial \bar{u}}{\partial t} - \frac{\partial^2 \bar{u}}{\partial y^2} = R_o \left[2 - \frac{\partial}{\partial y} \overline{uv} \right] ; \overline{uv} = -v_T \frac{\partial \bar{u}}{\partial y} \quad (1.20)$$

with the following given characteristic scales (2D channel width),

$$u_o = \left[-\frac{1}{\rho} \frac{dp}{dx} D \right]^{\frac{1}{2}} \quad (1.21)$$

and the initial and boundary conditions

$$\bar{u}(y, 0) = f_o(y) \quad (1.22)$$

$$\bar{u}(0, t) = \bar{u}(1, t) = 0 \quad (1.23)$$

R_0 stands for $u_0^2 2D/\nu$. Two kinds of viscosity models are retained leading to available predictions, namely

$$v_T = g(y) \int_0^1 \bar{u} dy \quad (1.24)$$

(model 1)

in which v_T is dimensionless parameter and

$$v_T(y, t) = R_0^{\frac{1}{2}} ky \sqrt{|\partial \bar{u}|}(0, t) [1 - \exp - \frac{y}{A} R_0] ; y \in [0; 0, 1] \quad (1.25)$$

$$v_T(y, t) = v_T(0, 1; t) ; y \in [0, 1; 0, 5] \quad (1.26)$$

$$v_T(y, t) = v_T(1 - y; t) ; y \in [0, 5, 1] . \quad (1.27)$$

(model 2)

Where second order closures are concerned such as previously proposed by Donaldson, Launder, Lumley, Jeandel... a partial differential system must be treated. The momentum equations and the continuity equation are supplemented by four equations related to the components of the Reynolds stress tensor. The dissipation rate must also be considered as an unknown function. Where a channel flow is concerned, the problem may be formulated in a rather simple manner. Using dimensionless variables for the Reynolds stresses and dissipation a model of this kind may be written in the general form:

$$R_0 \frac{\partial \vec{u}}{\partial t} - \frac{\partial^2 \vec{u}}{\partial y^2} = R_0 \vec{b} [\vec{u}] + R_0 \vec{b} \frac{\partial}{\partial y} \left[\frac{u_2 u_5}{u_3} \frac{\partial \vec{u}}{\partial y} \right] \quad (1.28)$$

U is the vector the components of which are:

$$u_1 = \bar{u} ; u_2 = \frac{\overline{u^2}}{u_0^2} ; u_3 = \bar{\varepsilon} \frac{2D}{u_0^3} ; u_4 = \frac{\overline{u_1^2}}{u_0^2} ; u_5 = \frac{\overline{u_2^2}}{u_0^2} ; u_6 = \frac{\overline{u_1 u_2}}{u_0^2} .$$

(1.29)

We must notice that inhomogeneities greatly complicate the modeling of the velocity pressure correlations. In the model of Launder,¹¹ a sophisticated

expression inversely proportional to the distance from the wall is incorporated in the non-linear operator $\bar{D} [\vec{U}]$. Consequently, boundary conditions for the mean and turbulent fields must be given in the well known semilogarithmic region.

We also notice that the Reynolds stress equations contain gradient terms in the form:

$$\frac{\partial}{\partial y} \left[\frac{U_2 U_5}{U_3} \frac{\partial \vec{U}}{\partial y} \right] \quad (1.30)$$

that reinforces the elliptic character of the equations governing the turbulent quantities.

A numerical solution may be searched by means of an optimal control formulation of the problem.¹² When non-linear operators are concerned, they can be split into two parts: a linear operator A and a non-linear one B; and we write:

$$A [U] = B[U] \quad . \quad (1.31)$$

The associated problem can be written:

$$A [U_\lambda] = B [\lambda] \quad . \quad (1.32)$$

In that case, λ may be considered as a control parameter, and the last state equation is solved minimizing a cost function of the form

$$J (\lambda) = \frac{1}{2} ||U_\lambda - \lambda||_v^2 \quad . \quad (1.33)$$

I.2.3 Provisional Comments

Even though these methods are carried out in physical space, references are often made to the size of the turbulent structures which are responsible for a phenomenon; in fact, an equilibrium state is always supposed which reinforces the role played by the energy flux which travels without alteration across the inertial range. If a deterministic phenomenon is superimposed on the random field (with vortex shedding, the spectrum exhibits a peak at a rather low wave number) the amount of energy flux can be significantly altered so that many of previous assumptions are spurious. Launder, Cousteix take into account this disturbance by introducing an additional parameter (as is made in the

thermodynamics of irreversible phenomena in order to account for a disequilibrium). It is also possible to directly work in spectral space.

1.3. Spectral Methods

Random motion is analysed accounting for the role played by turbulent structures which are referenced both by their sizes and orientations. Non-linear effects display an energy transfer; this energy flux generally travels from the largest to the smallest structures, but a reverse cascade process has been brought to light very clearly (Gence¹³). Starting from the Fourier transform of the two components u_i and u_j of the fluctuating velocity it is possible to write the equation of the correlation at two points \vec{k} and \vec{L} of spectral space (which corresponds to three points in physical space).

$$\begin{aligned} \left\{ \frac{\partial}{\partial t} + \gamma (k^2 + L^2) \right\} \langle \tilde{u}_i(\vec{k}, t) \tilde{u}_j(\vec{L}, t) \rangle = \\ + \frac{i}{2} P_{ilm}(\vec{k}) \int_{\vec{p}+\vec{q}=\vec{k}} \langle \tilde{u}_j(\vec{L}, t) \tilde{u}_l(\vec{p}, t) u_m(\vec{q}, t) \rangle d\vec{p} d\vec{q} \quad (1.34) \\ + \frac{i}{2} P_{jlm}(\vec{L}) \int_{\vec{p}'+\vec{q}'=\vec{L}} \langle \tilde{u}_i(\vec{k}, t) u_l(\vec{p}', t) u_m(\vec{q}', t) \rangle d\vec{p}' d\vec{q}' \end{aligned}$$

Integrating over the wave number \vec{L} it is possible to find the classical spectral equation at a unique point \vec{k} of spectral space. The general structure of this equation can be represented by

$$\frac{\partial}{\partial t} \langle \tilde{u} \tilde{u} \rangle = \langle \tilde{u} \tilde{u} \tilde{u} \rangle \quad (1.35)$$

For this open problem it is also possible to write an equation for the triple correlations at three points in spectral space

$$\begin{aligned} \left\{ \frac{\partial}{\partial t} + \nu(k^2 + p^2 + q^2) \right\} \langle u_i(\vec{k}, t) u_l(\vec{p}, t) u_m(\vec{q}, t) \rangle = \\ - \frac{i}{2} P_{irs}(\vec{k}) \int_{i+j=k} \langle u_l(\vec{p}, t) u_m(\vec{q}, t) u_r(\vec{r}, t) u_s(\vec{j}, t) \rangle d\vec{i} d\vec{j} \\ + \dots + \dots = S_{ilm}(\vec{p}, \vec{q}, \vec{k}, t) \quad (1.36) \end{aligned}$$

To close this open set of equations Millionshikov (1941) and Proudman and Reid (1954)¹⁴ assumed that the probability law

$$P [\vec{u}_i, \vec{u}_l, \vec{u}_r, \vec{u}_s] \quad (1.37)$$

which is parametrized by $\vec{k}, \vec{p}, \vec{l}, \vec{q}$ and t is very similar to a gaussian law so that they write:

$$\langle \tilde{u}_i \tilde{u}_l \tilde{u}_r \tilde{u}_s \rangle = \langle \tilde{u}_i \tilde{u}_l \rangle \langle \tilde{u}_r \tilde{u}_s \rangle + \langle \tilde{u}_i \tilde{u}_r \rangle \langle \tilde{u}_l \tilde{u}_s \rangle + \langle \tilde{u}_i \tilde{u}_s \rangle \langle \tilde{u}_l \tilde{u}_r \rangle \quad (1.38)$$

This assumption leads to:

$$\langle \tilde{u}_i \tilde{u}_l \tilde{u}_r \rangle = 0 \quad (1.39)$$

Disregarding some contradictions these authors admit that a closure method can be given in the symbolic form:

$$\frac{\partial}{\partial t} \langle u u u \rangle = \langle u u \rangle \langle u u \rangle \quad (1.40)$$

Thereby, the role played by the cumulant terms is neglected. In the previous equation no limitation is imposed on the evolution of the triple correlations which increase too much. These terms being responsible for the energy transfer across the spectrum, the amount of energy located in some regions becomes negative which is irrelevant. With the quasi normal hypothesis the solution of the equation for the triple correlations

$$\left\{ \frac{\partial}{\partial t} + \gamma (k^2 + p^2 + q^2) \right\} \langle \tilde{u}_i(\vec{k}, t) u_l(\vec{p}, t) u_m(\vec{q}, t) \rangle = s_{ilm}[\vec{k}, \vec{p}, \vec{q}, t] \quad (1.41)$$

is given by

$$\begin{aligned} \langle \tilde{u}_i(k, t) \tilde{u}_l(p, t) \tilde{u}_m(q, t) \rangle &= \langle \tilde{u}_i \tilde{u}_l \tilde{u}_m \rangle e^{-\nu(k^2 + p^2 + q^2)t} \\ &+ \int_0^t s_{ilm}[k, p, q, t'] e^{-\nu(k^2 + p^2 + q^2)(t-t')} dt', \end{aligned} \quad (1.42)$$

where S_{ilm} must be considered as a known quantity. It is obvious that the memory of the turbulence is overestimated, the exponential term being only controlled by the kinematic viscosity. In order to improve the method, Orszag¹⁵ supposed that the role played by the cumulant terms can be introduced through a linear relaxation which has the symbolic form

$$\frac{\partial}{\partial t} \langle u u u \rangle = \langle u u \rangle \langle u u \rangle - \mu_T \langle u u u \rangle . \quad (1.43)$$

This artificial viscosity contributes to reduce the memory time of turbulence. Moreover the integration is drastically simplified by supposing that the triple correlations can be taken at time t (in place of t'). From this Markovian approximation the integration with respect to t is only carried out over the exponential term. The artificial viscosity μ_T plays a role in spectral space which could be compared to the turbulent viscosity introduced by Boussinesq in physical space. Starting from experimental data, adjustments can be made, but it is also possible to evaluate the characteristic time by considering, in the manner of Kraichnan, a supplementary problem which is the advection of a passive scalar by a random field with coupled variables (Test Field Model TFM).

The eddy damped quasi normal markovian model (E.D.Q.N.M.) has been extended to homogeneous fields with constant mean velocity gradients by Cambon.¹⁶ The spectral equation has previously been given by Mitchner and Craya:

$$\left\{ \frac{\partial}{\partial t} + v (k^2 + p^2 + q^2) \right\} \phi_{lji} [\vec{k}, \vec{p}, t] + \psi_{lji} [\vec{k}, \vec{p}, t] = \Omega_{lji} [\vec{k}, \vec{p}, t] \quad (1.44)$$

with

$$\vec{p} + \vec{q} = \vec{k} . \quad (1.45)$$

The action of the mean velocity gradient is explicitly represented by ψ_{lji} whereas Ω_{lji} accounts for non-linear effects. We suppose Ω_{lji} to be evaluated in the form:

$$\Omega_{lji} [\vec{k}, \vec{p}] = \Omega_{lji}^{QN} [\vec{k}, \vec{p}] - L_{ljimnr}^{ED} \phi_{mnr} [\vec{k}, \vec{p}] . \quad (1.46)$$

We find that

$$\left\{ \frac{\partial}{\partial t} + v (k^2 + p^2 + q^2) \right\} \phi_{lji} [\vec{k}, \vec{p}, t] + \psi_{lji} [\vec{k}, \vec{p}, t] = \quad (1.47)$$

$$\Omega_{lji}^{QN} - L_{ljimnr} \phi_{mnr} \quad .$$

Besides it is written:

$$v(k^2 + p^2 + q^2) \phi_{lji} + \psi_{lji} + L_{ljimnr}^{ED} \phi_{mnr} = \mu_T \phi_{lji} \quad . \quad (1.48)$$

The incorporation of the typically non-isotropic term ψ_{lji} is a diagonalized form has extensively been discussed. Two facts can support this assumption. At first, the variation of ϕ_{lji} with time should be specially controlled by Ω_{lji}^{QN} . Secondly we are only concerned with a part of the triple correlations for the behavior of the equations which determine the evolution of the double correlations.²

In order to simplify computing methods, the evolution of the turbulent structures is only considered as a function of their sizes, not of their orientation. Consequently an integration is made over a sphere of radius $|\vec{k}|$ so that suitable functions are introduced such as:

$$\phi_{ij} [|\vec{k}|, t] = \int_S \phi_{ij} [\vec{k}, t] d\Sigma(\vec{k}) \quad . \quad (1.49)$$

An equation in terms of functions of modulus of \vec{k} is obtained. An angular parametrization of the second-order spectral tensor is introduced in order to integrate analytically all the directional terms over a spherical shell. Modelling methods similar to those used in physical space are introduced; for example, the spectral deviatoric tensor H_{ij} plays a similar role to b_{ij} . Finally, the terms of the closed equation for ϕ_{ij} exhibit only a tensorial anisotropy. The specific anisotropy related to directional effects disappears by integrating over the sphere. At this stage, two scalar parameters $a(k, t)$ and $\eta(k, t)$ are to be adjusted. Exact linear solutions can successfully be used to determine $a(k, t)$. Complex evolution of spectral curves can be predicted, for instance a peak can be introduced in a local range of wave number and direct and reversed cascades of energy can be brought to light (Fig. 1 and 2).

Many extensions of this method have been made to turbulent fields including buoyancy effects, inertial forces and so on.

I.4. Pseudo Deterministic Approach (large eddy simulation)

Starting from the Navier-Stokes equations, direct simulations of isotropic turbulent fields have been made. So far, this simulation is not adequately supported by mathematical considerations. Bifurcation points, if they exist, are detected and treated through an algorithmic scheme which should introduce its own mark so that it should be difficult to interpret the final results which include both the properties of the initial equations and those of the numerical scheme.

We can suppose that a large eddy simulation is more typically convenient for turbulent fields subjected to overall effects capable of influencing more or less extended groups of turbulent structures. Except for isotropic turbulence which is only subjected to its own mechanism, in most of the flow patterns, selected turbulent structures are subjected to straining mechanisms or driving forces such as mean velocity gradients, buoyancy effects, etc. Therefore, it is possible to conjecture the existence of coherent structures whose development could be more suitably tackled in a deterministic way. These coherent structures, the sizes of which are comparable with those of the overall flow, develop in the presence of smaller structures which are accounted for in the statistical coefficients. For example, the pairing process which is responsible to a large extent for the growth of turbulent areas can be considered as a deterministic mechanism whose developments are somewhat hampered by the fine grained turbulence which plays a role similar to a viscous effect.

So far, the definition of such coherent structures is not precisely supported by mathematical theories. In these approaches, Deardorff¹⁷ and Leonard¹⁸ introduce a typical splitting of the overall field. The "filter function" used by Leonard can be linked to the spectral approach more directly. Therefore we use it preferentially.

Leonard separates the velocity field $U_i(\vec{x}, t)$ into a grid scale component $\bar{u}_i(\vec{x}, t)$ and a subgrid component $u'_i(\vec{x}, t)$

$$\bar{u}_i(\vec{x}, t) = \int G[\vec{x}/\vec{x}'] U_i(\vec{x}', t) d\vec{x}'$$

$$U_i = \bar{u}_i + u'_i \quad ; \quad \int G[\vec{x}/\vec{x}'] d\vec{x}' = 1 \quad . \quad (1.50)$$

The overbar precisely denotes the filtering operation which is not to be confused with either the mean value of the velocity, which has a statistical

meaning, or with any statistical averaging, indicated by angular brackets. A filtered quantity has to be considered as a typical event which can occur under any circumstance. This event, when it happens, develops in an environment the properties of which have to be considered in connection with the existence of this principal event. This should lead to conditional samplings supplemented by averaging operations.

If the physical properties of the fluid are constant, the result of filtering the Navier-Stokes equation is:

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial}{\partial x_j} \overline{u_i u_j} = - \frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_i} + \nu \nabla^2 \bar{u}_i . \quad (1.51)$$

This equation is supplemented by the continuity equation:

$$\frac{\partial \bar{u}_j}{\partial x_j} = 0 . \quad (1.52)$$

For the advection term it is found that:

$$\overline{u_i u_j} = \bar{u}_i \bar{u}_j + \overline{u'_i \bar{u}_j} + \overline{\bar{u}_i u'_j} + \overline{u'_i u'_j} . \quad (1.53)$$

By using a series expansion it is found that:

$$\frac{\partial}{\partial x_j} \overline{u_i u_j} = \frac{\partial}{\partial x_j} [\bar{u}_i \bar{u}_j] + \frac{\partial}{\partial x_j} \left[\frac{\Delta^2}{\gamma} \frac{\partial}{\partial x_k} \frac{\partial}{\partial x_k} \bar{u}_i \bar{u}_j \right] . \quad (1.54)$$

The Leonard term can be written

$$L_{ij} = \overline{\bar{u}_i \bar{u}_j} - \bar{u}_i \bar{u}_j . \quad (1.55)$$

The product of filtered quantities is considered without additional filtering.

The "cross term"

$$C_{ij} = \overline{\bar{u}_i u'_j} + \overline{u'_i \bar{u}_j}$$

represents interactions between the residual field and the filtered field. "A priori" this sort of interaction should not be located in spectral space.

On the other hand, the subgrid Reynolds tensor

$$R_{ij} = \overline{u'_i u'_j}$$

emphasizes the role of small structures. However, it can be proved that such interactions can occasionally generate structures far outside the subgrid scale from a triadic analysis.

We set

$$T_{ij} = R_{ij} + C_{ij}$$

and introduce the deviatoric tensor:

$$\tau_{ij} = T_{ij} - \frac{1}{3} T_{kk} \delta_{ij} \quad (1.56)$$

On the other hand we write:

$$\hat{\bar{p}} = \frac{\bar{p}}{\rho} + \frac{1}{3} T_{kk} \quad (1.57)$$

Finally we have:

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial}{\partial x_j} \bar{u}_i \bar{u}_j = - \frac{1}{\rho} \frac{\partial \hat{\bar{p}}}{\partial x} - \frac{\partial}{\partial x_j} [L_{ij} + \tau_{ij}] \quad (1.58)$$

Subsequent workers have followed Smagorinsky¹⁹ and write

$$\tau_{ij} = - \nu_N \left[\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right] \quad (1.59)$$

for the "subgrid term".

Such an assumption gives rise to several remarks:

No tractable information has first been extracted for the subgrid term from the basic equations.

With such a splitting, the subgrid terms are not objective terms. Therefore, the application of representative theorems to these quantities becomes questionable.*

For a long time the energy cascade was considered as a one way street, the small structures being fed by the larger ones. Moreover it was conjectured that interactions were only possible between turbulent eddies having approximately the same sizes. This opinion was more or less supported by experimental data extracted from isotropic turbulent fields. In fact, the energy cascade is a two way street and interaction occurs between structures of very different sizes.

Applications of successive plane strains to grid generated turbulence has been illuminating (Gence 13-20). At first the isotropic turbulence is subjected to a straining process and turbulent structures are oriented by the mean field. A second straining, the axes of which are different from the first ones by an angle α is suddenly applied so that the turbulence is reoriented by this second straining. Gence¹³ has shown that the term describing the exchange of energy can be written, after the change in position of the principal axes of the strain, as

$$-D_q^2 [b_{22} - b_{33}][2 \cos^2 \alpha - 1] . \quad (1.60)$$

Either positive or negative values can be obtained depending on the angle α , D stands for $\partial \bar{u}_i / \partial x_j$.

The assumption of Smagorinsky supposes the principal axes of the subgrid turbulence to be the same as those of the filtered quantities. This hypothesis is not supported by Gence's experiments and by direct simulations carried out by Ferziger²¹ who states that "the two sets of axes appear to be essentially random with respect to each other." The mean relative position of the two systems is of course a most important parameter. Occasionally such a coincidence could occur.

* At first, representation theorems require the introduction of additional argument in connection with, for instance, the rotation of the filtered field.

Comparisons have been made between the computational results coming from an extended E.D.Q.N.M. method (Cambon¹⁶) and from large eddy simulations including Smagorinsky's assumption (Courty, Bertoglio, private communication), (Fig. 3).

Great discrepancies are shown in Fig. 3, probably originating from the subgrid model. With the large eddy simulation, the evolution of the correlation term seems to be rather slow, a fact which should be in close connection with the evolution of the transverse velocity component.

1.5 Probabilistic Approach

So far, the probabilistic properties of the flow have been treated disregarding the precise behaviour of the Navier-Stokes equation which controls any motion of the complex turbulent medium. The motion is split into two parts, the properties of the fluctuating part are supposed not to be far from a normal law. At any point in the theory, it is necessary to clarify whether bifurcation points exist or do not. In fact, all the pervious treatments are supported by very pragmatic considerations not typically related to the nature of the basic equation. However, large eddy simulations start from the Navier-Stokes equation for a selected group of structures, but the deterministic computation is carried out up to an arbitrary solution no averaging process is introduced in order to give a statistical value to this approach. Besides the evaluation of the subgrid terms is very rough.

Simple experience confirms that some statistical properties of turbulent fields can behave in a very typical way; experimental data show that the departure from a normal law can be very significant, so that it is not possible to account for the cumulant terms through a linear relaxation as is done in the case of the E.D.Q.N.M. assumption. Turbulent fields with chemical reactions display such typical behaviour. If two non-reactive chemical species are mixed together, it is found that the statistical properties of the concentration fluctuations more or less relax to a gaussian curve, but if the two species exhibit very strong reactive properties A and B cannot coexist at the same point, so that the probability density function for the two fluctuating concentrations should be represented by two planes. Even though asymptotic situations such as flames could be easier to tackle, it would be interesting to follow the evolution of the probability density function by controlling the reaction rate. Experimental investigations carried out by Bennanni²² are in progress to seize intermediate situations. The Damkohler Number compares two characteristic scales: the reaction time, which is directly connected with chemical properties, and a characteristic time linked

to dynamical properties of the turbulent field such as $\overline{q^2}/\varepsilon$. At first, it seems that the Damkholer Number does not act on the probability density function in a continuous way so that it should be possible to encounter a probability distribution not far from a Gaussian curve for large ranges of small chemical reaction rate. It is probably interesting to seize the topology of the reactive domain. For flames where the rate of reaction is high, reactive zones are complex surfaces; where diluted media are concerned, reactive zones can become more or less extended volumes.

For the treatment of chemical reactions the trend should be to consider a probability density equation, a means which has recently been used.

In the cases under consideration, the theory of moments having failed, scientists have introduced a fine probability function. The solutions have been developed in very simple cases. For our purpose, we must underline the role played by the Hopf equation which can be considered as the starting point for this theory as it has been shown by Gence (private communication). The Hopf equation can be considered as a probabilistic form of the Navier-Stokes equation. It is established in phase space and supposes the uniqueness of the solution when convenient initial conditions are given. This restriction is introduced through a punctual correspondence between the sets of points which represent the state of the system at two selected times. In the original paper by Hopf, this difficulty is perfectly brought to light.

For the sake of simplicity, we consider an isotropic turbulence which develops in a reactive medium, and we indicate the equation governing the probability density function:

$$\varepsilon^1 [\Gamma_A^1, \Gamma_B^1, t] \quad (1.61)$$

so that $\varepsilon [\Gamma_A^1, \Gamma_B^1, t] d\Gamma_A^1 d\Gamma_B^1$ represents the probability of having the concentration $c_A(\vec{x}_1, t)$ and $c_B(\vec{x}_1, t)$ respectively located in a range $\Gamma_A^1, \Gamma_A^1 + d\Gamma_A^1$ and $\Gamma_B^1, \Gamma_B^1 + d\Gamma_B^1$ at a given time t . Finally we obtain:

$$\begin{aligned} \frac{\partial \varepsilon^1}{\partial t} = & - \frac{\partial}{\partial \Gamma_A^1} \{ \gamma(\Gamma_A^1, \Gamma_B^1) \varepsilon^1 \} - \frac{\partial}{\partial \Gamma_B^1} \{ \gamma(\Gamma_A^1, \Gamma_B^1) \varepsilon^1 \} \\ & = \lim_{t \rightarrow 0} \Delta \{ \alpha_A \frac{\partial}{\partial \Gamma_A^1} \int \Gamma_A^2 \varepsilon^2 d\Gamma_A^2 d\Gamma_B^2 + \alpha_B \frac{\partial}{\partial \Gamma_B^1} \int \Gamma_B^2 \varepsilon^2 d\Gamma_A^2 d\Gamma_B^2 \} \end{aligned} \quad (1.62)$$

where

$$\varepsilon^2 = \varepsilon^2 [\Gamma_A^1, \Gamma_B^1, \Gamma_A^2, \Gamma_B^2, \vec{r}, t] \quad (1.63)$$

is the probability density of finding given values of concentration at two separate points \vec{r} . In this form, the closure problem is clearly brought to light. At the present time, this approach is an unweeded path. Experimental data are unavailable and information is required to suggest suitable closures.

II. MATHEMATICAL STUDY OF A TURBULENT VISCOSITY MODEL

II.1. Introduction

In this part, we shall present a mathematical study of a nonlinear parabolic equation including one of the above turbulent viscosity models. We wish to focus our attention on the more original one, namely model 2 (see Section I.2.2). For a mathematical study of model 1, see Lainé.^{23,24} In this paper, the proofs will be only outlined. For further details, we refer the reader to Lainé,²³ Brauner and Lainé.²⁵

Our mathematical notations will be as follows: Ω is the open interval $]0,1[$, x the space variable. Let $]0,S[$ be the time interval, t or s the time variable, and $Q = \Omega \times]0,S[$. We shall use some Sobolev spaces (see e.g., Adams²⁶), namely $H_0^1(\Omega)$, $H^2(\Omega)$, $H^{-1}(\Omega)$ (³), as well as the spaces $L^2(0,S; H_0^1(\Omega))$, etc. Without any further specification, we shall denote by $(,)$ the inner product in $L^2(\Omega)$ and $|| \cdot ||$ the associated norm.

For u regular enough, let us define the "turbulent viscosity"

$$v_T(u) = G \sqrt{\left| \frac{\partial u}{\partial x}(0,t) \right|} \quad (2.1)$$

where $G \in C^1(\bar{\Omega})$, $G(x) \geq v_1 > 0$.

We consider the following mathematical problem. Find $u = u(x,t)$ solution of

$$\begin{cases} \frac{\partial u}{\partial t} - v_0 \frac{\partial^2 u}{\partial x^2} - \frac{\partial}{\partial x} (v_T(u) \frac{\partial u}{\partial x}) = f \\ u(0,t) = u(1,t) = 0 \\ u(x,0) = u_0(x), \quad x \in \Omega \end{cases} \quad (2.2)$$

where $v_0 \gg v_1$ and $v_T(u)$ is given by (2.1).

II.2. Existence of a Solution

The existence of a solution to problem (2.2) is obtained via a standard Galerkin method. Specifically, we establish the following theorem:

Theorem 2.1: Let f given in $L^2(Q)$, u_0 in $H_0^1(\Omega)$. Then problem (2.2) has a solution u in the space $L^2(0, S; H^2(\Omega)) \cap L^\infty(0, S; H_0^1(\Omega))$.

Furthermore,

$$\left| \frac{\partial u}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial u}{\partial x} \text{ and } \left| \frac{\partial u}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial^2 u}{\partial x^2} \text{ belong to } L^2(Q). \quad (2.3)$$

Proof: Introduce the eigenfunctions w_j of the operator $-d^2/dx^2$, namely $w_j(x) = \sin j\pi x$, $j = 1, 2, \dots$. One has $w_j'' + \lambda_j w_j = 0$ where $\lambda_j = (j\pi)^2$. We look for a solution $u_m(t) = \sum_{j=1}^m g_{jm}(t) w_j$ of the nonlinear differential system:

$$\left\{ \begin{aligned} & \left(\frac{\partial u_m}{\partial t}(t), w_j \right) - v_0 \left(\frac{\partial^2 u_m}{\partial x^2}(t), w_j \right) - \left(\frac{\partial}{\partial x} (v_T(u_m)) \frac{\partial u_m}{\partial x}(t), w_j \right) \\ & \qquad \qquad \qquad = (f(t), w_j), \quad j = 1, \dots, m; \\ & u_m(0) = u_{0m} = \text{proj. of } u_0 \text{ on } V_m \end{aligned} \right. \quad (2.4)$$

System (2.4) has a solution on an interval $]0, S_m[$. The following estimates will in particular yield that $S_m = S$.

Lemma 2.1: The sequence $\{u_m\}$ is bounded in the space $L^2(0, S; H_0^1(\Omega))$

$\cap L^\infty(0, S; L^2(\Omega))$, and

$$\left| \frac{\partial u_m}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial u_m}{\partial x} \text{ is bounded in } L^2(Q).$$

Proof: Simply multiply (2.4) by $g_{jm}(t)$ and add for $j = 1$ to m . It follows

$$\frac{1}{2} \frac{d}{dt} \|u_m(t)\|^2 + v_0 \left\| \frac{\partial u_m}{\partial x}(t) \right\|^2 + \int_{\Omega} v_T(u_m) \left(\frac{\partial u_m}{\partial x} \right)^2 dx = (f(t), u_m(t))$$

The integrate over $]0, S[$ and apply Young's formula to the R.H.S.

Lemma 2.2: The sequence $\{u_m\}$ is bounded in the space $L^2(0, S; H^2(\Omega)) \cap L^\infty(0, S; H_0^1(\Omega))$, and

$$\left| \frac{\partial u_m}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial^2 u_m}{\partial x^2} \text{ is bounded in } L^2(Q).$$

Sketch of the proof: In (2.4), replace w_j by $-1/\lambda_j$, then multiply by $g_{jm}(t)$ and sum. It comes:

$$\begin{aligned} & - \left(\frac{\partial u_m}{\partial x}(t), \frac{\partial^2 u_m}{\partial x^2}(t) \right) + v_0 \left\| \frac{\partial^2 u_m}{\partial x^2}(t) \right\|^2 + \int_{\Omega} v_T(u_m) \left(\frac{\partial^2 u_m}{\partial x^2}(t) \right)^2 dx \\ & + \left| \frac{\partial u_m}{\partial x}(0, t) \right|^{\frac{1}{2}} \int_{\Omega} g' \frac{\partial u_m}{\partial x}(t) \frac{\partial^2 u_m}{\partial x^2}(t) dx = - \left(f(t), \frac{\partial^2 u_m}{\partial x^2}(t) \right) \end{aligned}$$

Integrating over $]0, S[$ leads to

$$\begin{aligned} & \left\| \frac{\partial u_m}{\partial x}(s) \right\|^2 + 2v_0 \left\| \frac{\partial^2 u_m}{\partial x^2} \right\|_{L^2(Q)}^2 + 2 \int_Q v_T(u_m) \left(\frac{\partial^2 u_m}{\partial x^2} \right)^2 dx dt \\ & = \left\| \frac{\partial u_m}{\partial x} \right\|^2 + 2 \int_Q f \frac{\partial^2 u_m}{\partial x^2} dx dt - 2 \int_Q \left| \frac{\partial u_m}{\partial x}(0, t) \right|^{\frac{1}{2}} g' \frac{\partial u_m}{\partial x} \frac{\partial^2 u_m}{\partial x^2} dx dt. \end{aligned}$$

The last term in the R.H.S. may be bounded by

$$2 \|g'\|_{L^\infty(\Omega)} \left\| \left| \frac{\partial u_m}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial u_m}{\partial x} \right\|_{L^2(Q)} \cdot \left\| \left| \frac{\partial u_m}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial^2 u_m}{\partial x^2} \right\|_{L^2(Q)}$$

which can be split into two parts using Young's formula. Finally we get:

$$\begin{aligned} & \left\| \frac{\partial u_m}{\partial x}(s) \right\|^2 + v_0 \left\| \frac{\partial^2 u_m}{\partial x^2} \right\|_{L^2(Q)}^2 + v_1 \left\| \left| \frac{\partial u_m}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial^2 u_m}{\partial x^2} \right\|_{L^2(Q)}^2 \\ & \leq \frac{1}{v_0} \|f\|_{L^2(Q)}^2 + \left\| \frac{\partial u_0}{\partial x} \right\|^2 + \frac{1}{v_1} \|g'\|_{L^\infty(\Omega)} \cdot \left\| \left| \frac{\partial u_m}{\partial x}(0, \cdot) \right|^{\frac{1}{2}} \frac{\partial u_m}{\partial x} \right\|_{L^2(Q)} \end{aligned}$$

hence the requested estimates (for the estimate in $L^\infty(0, S; H_0^1(\Omega))$), just integrate over $]0, s[$, $0 < s < S$. ■

End of the proof of Theorem 2.1: As system (2.4) may be rewritten as

$$\frac{\partial u_m}{\partial t} = \nu_0 P_m \left(\frac{\partial^2 u_m}{\partial x^2} \right) + P_m \left(\frac{\partial}{\partial x} \nu_T(u_m) \frac{\partial u_m}{\partial x} \right) + P_m f, \quad (2.5)$$

where P_m is the projection operator from $L^2(\Omega)$ into V_m , it follows that the sequence

$$\left\{ \frac{\partial u_m}{\partial t} \right\}$$

is bounded in $L^2(0, T; H^{-1}(\Omega))$. Hence we can extract a converging subsequence $\{u_\mu\}$ such that $u_\mu \rightharpoonup u$ in $L^2(0, S; H^2(\Omega))$ weak and in $L^\infty(0, S; H_0^1(\Omega))$ weak star,

$$\frac{\partial u_\mu}{\partial t} \rightharpoonup \frac{\partial u}{\partial t}$$

in $L^2(0, S; H^{-1}(\Omega))$ weak. By a compactness theorem (see Lions [27]), $u_\mu \rightarrow u$ in $L^2(0, S; H^{2-\varepsilon}(\Omega))$ strong, $0 < \varepsilon < \frac{1}{2}$, and, by Sobolev imbedding, in $L^2(0, S; C^1(\bar{\Omega}))$. Then it is easy to pass to the limit in (2.4), to see that μ is solution of (2.2).

II.3. Results on Regularity

In order to prove the uniqueness, we must first establish further results on regularity, namely:

Theorem 2.2: Suppose now $f \in L^2(0, S; H_0^1(\Omega))$, $G \in C^2(\bar{\Omega})$, $G'(0) = G'(1) = 0$,

$G(x) \geq \nu_1 > 0$, and $u_0 \in H^2(\Omega) \cap H_0^1(\Omega)$. Then problem (2.2) has a solution in $L^2(0, S; H^3(\Omega)) \cap L^\infty(0, S; H^2(\Omega))$, $\frac{\partial u}{\partial t} \in L^2(0, S; H^1(\Omega))$. Besides:

$$\left| \frac{\partial u}{\partial x}(0, \cdot) \right|^4 \frac{\partial^3 u}{\partial x^3} \in L^2(Q). \quad (2.6)$$

Sketch of the proof: In (2.4), replace w_j by $1/\lambda_j^2 \frac{d^4 w_j}{dx^4}$:

$$\begin{aligned}
& \left(\frac{\partial u_m}{\partial t} (t), \frac{\partial^4 u_m}{\partial x^4} (t) - v_0 \left(\frac{\partial^2 u_m}{\partial x^2} (t), \frac{\partial^4 u_m}{\partial x^4} (t) \right) \right. \\
& \left. - \left(\frac{\partial}{\partial x} (v_T (u_m)) \frac{\partial u_m}{\partial x} (t), \frac{\partial^4 u_m}{\partial x^4} (t) \right) = (f(t), \frac{\partial^4 u_m}{\partial x^4} (t)) \right).
\end{aligned}$$

Integrating by parts gives:

$$\begin{aligned}
& \frac{1}{2} \frac{d}{dt} \left\| \frac{\partial^2 u_m}{\partial x^2} (t) \right\|^2 + v_0 \left\| \frac{\partial^3 u_m}{\partial x^3} (t) \right\|^2 \\
& = - \left(\frac{\partial f}{\partial x} (t), \frac{\partial^3 u_m}{\partial x^3} (t) \right) - \left(\frac{\partial^2}{\partial x^2} (v_T (u_m)) \frac{\partial u_m}{\partial x} (t), \frac{\partial^3 u_m}{\partial x^3} (t) \right).
\end{aligned}$$

We refer to Brauner-Lainé (1981) for the estimate of the R.H.S.

Corollary 2.1. u is continuous in $\bar{\Omega}$ and $\frac{\partial u}{\partial x} (0, \cdot) \in C^0([0, S])$.

Proof: From Theorem 2.2, we know $u \in L^2(0, S; H^3(\Omega))$, $\frac{\partial u}{\partial t} \in L^2(0, S; H^1(\Omega))$. By an interpolation argument (Lions-Magenes [28]), $u \in C^0([0, S]; H^2(\Omega))$, and by Sobolev imbedding, $u \in C^0([0, S]; C^1(\bar{\Omega}))$.

II.4. A Strong Maximum Principle for Weak Solutions

In as much as (2.2) is a physical problem, we are only concerned in positive data. What we need to prove is that any solution obtained via Theorem 2.1 and 2.2 is not only non-negative, but positive in Q . A consequence, we will establish $\frac{\partial u}{\partial x} (0, t) > 0$.

Theorem 2.3: The hypotheses are that of Theorem 3.2. Besides suppose

$$\begin{aligned}
& u_0(x) > 0 \quad \forall x \in \Omega, \quad f(x, t) \geq 0 \text{ a.e. in } Q. \text{ Then solution of (2.2)} \\
& \text{in the sense of Theorem 2.2 verifies } u(x, t) > 0 \text{ in } Q.
\end{aligned}$$

First it is easy to check $u(x, t) \geq 0$ with the weak maximum principle. Then $v_T(u) = \frac{\partial u}{\partial x} (0, t)$. Let us now mention a result due to J. Moser [29, 30].

II.4.1 A Harnack Inequality for Parabolic Equations

Let Ω denote an open bounded set in \mathbb{R}^n and $Q = \Omega \times]0, T[$. We consider the equation in Q :

$$\frac{\partial v}{\partial t} - \sum_{i,j} \frac{\partial}{\partial x_j} (a_{ij} (x, t) \frac{\partial v}{\partial x_i}) = 0 \tag{2.7}$$

where a_{ij} are bounded functions such that

$$\begin{cases} 0 < \lambda \leq \sum_{i,j} a_{ij} \xi_i \xi_j < \Lambda < +\infty, \quad \forall \xi = (\xi_i) \in \mathbb{R}^n, \quad |\xi|^2 = 1 \\ a_{ij} = a_{ji} \end{cases} \quad (2.8)$$

Let v be a weak solution of (2.7) such that

$$\frac{\partial u}{\partial t}, \quad \frac{\partial u}{\partial x_i}, \quad i = 1, \dots, n,$$

belong to $L^2(Q)$. We suppose that v is non-negative in Q . We consider a compact and connected subdomain of Ω , A , and two subintervals $I^- = \{t, t_1 < t < t_2\}$, $I^+ = \{t, t_3 < t < t_4\}$, where we assume that $0 < t_1 < t_2 < t_3 < t_4 < T$. Define $Q^- = I^- \times A$, $Q^+ = I^+ \times A$.

Lemma 2.3: Any non-negative solution of (2.7) satisfies:

$$\sup_{Q^-} v \leq c^{(1 + \lambda^{-1})} \inf_{Q^+} v \quad (2.9)$$

where $c > 1$ is a constant which depends only on Q , Q^- and Q^+ .

II.4.2 Positivity of u

Now we can prove Theorem 2.3. Let $u \geq 0$ a solution of (2.2), and consider the following auxiliary problem:

$$\begin{aligned} \frac{\partial \tilde{u}}{\partial t} - v_0 \frac{\partial^2 \tilde{u}}{\partial x^2} - \left(\frac{\partial u}{\partial x}(0, t) \right)^{\frac{1}{2}} \frac{\partial}{\partial x} \left(G \frac{\partial \tilde{u}}{\partial x} \right) &= 0 \\ \tilde{u}(0, t) = \tilde{u}(1, t) &= 0 \\ \tilde{u}(x, 0) &= u_0(x) \end{aligned} \quad (2.10)$$

Lemma 2.4: $\tilde{u} \in C^0(\bar{Q})$ and $u(x, t) \geq \tilde{u}(x, t) > 0$ in Q .

Proof: It is easy to verify $u \geq \tilde{u} \geq 0$ by the weak maximum principle. The main point is to prove $\tilde{u} > 0$. As the 2nd member in (2.10) is zero, we can

apply Lemma 2.3 to (2.10), with $a_{ij} = a = v_T(u) + v_0$. Here $\lambda = v_0$ and

$$\Lambda = v_0 + \left\{ \sup_{\bar{\Omega}} G \cdot \sup_{[0, S]} \left(\frac{\partial u}{\partial x}(0, t) \right)^{\frac{1}{2}} \right\}.$$

Now let $x_0 \in \Omega$ fixed. Since $u_0 > 0$ in Ω , there exist a closed interval $A \subset \Omega$ containing x_0 , and a time τ , such that

$$\tilde{u}(x, t) > 0 \quad \forall x \in A, \quad 0 \leq t \leq \tau. \quad (2.11)$$

Suppose there exists $t_* > \tau$ such that $\tilde{u}(x_0, t_*) = 0$. With the notations of Lemma 2.3, we choose t_1 and t_2 such that $0 < t_1 < t_2 < \tau$, then we take $t_3 = t_* - \varepsilon$, $t_4 = t_* + \varepsilon$. So

$$\inf_{Q^+} \tilde{u} = 0, \text{ therefore } \sup_{Q^-} \tilde{u} = 0$$

as a result of (2.9). Of course this is inconsistent with (2.11) and t_* does not exist.

II.4.3 Positivity of $\frac{\partial u}{\partial x}(0, t)$

Lemma 2.5: $\frac{\partial u}{\partial x}(0, t) > 0 \quad \forall t \in]0, S[.$

The idea of the proof is to extend a demonstration by Protter-Weinberger [31] to positive weak solutions. See the details in [23][25].

Corollary 2.2: Suppose

$$\frac{du}{dx}(0) > 0.$$

Then exists a constant $\alpha > 0$ such that

$$\frac{\partial u}{\partial x}(0, t) \geq \alpha > 0 \quad \forall t \in [0, S]. \quad (2.12)$$

II.5 Uniqueness of the Solution

The demonstration of the uniqueness is based on Gronwall's lemma, with Corollary 2.2 as basic tool.

Theorem 2.4: The hypotheses are that of Theorem 2.3. Then problem (2.2) has a unique positive solution.

Sketch of the proof: Let u_1 and u_2 be two positive solutions of (2.2). Set $u = u_1 - u_2$. Clearly, u verifies

$$\frac{\partial u}{\partial t} - v_0 \frac{\partial^2 u}{\partial x^2} - \frac{\partial}{\partial x} (v_T(u_1) \frac{\partial u}{\partial x}) = \frac{\partial u}{\partial x} (v_T(u_1) - v_T(u_2)) \frac{\partial u_2}{\partial x}$$

$$u(0, t) = u(1, t) = 0 \quad (2.13)$$

$$u(x, 0) = 0.$$

Multiply (2.13) by $-\frac{\partial^2 u}{\partial x^2}$:

$$\left\{ \begin{aligned} & \frac{1}{2} \frac{d}{dt} \left\| \frac{\partial u}{\partial x}(t) \right\|^2 + v_0 \left\| \frac{\partial^2 u}{\partial x^2}(t) \right\|^2 + \int_{\Omega} v_T(u_1) \left(\frac{\partial^2 u}{\partial x^2} \right)^2 dx = \\ & - \int_{\Omega} \left(\frac{\partial u_1}{\partial x}(0, t) \right)^{\frac{1}{2}} G' \frac{\partial u}{\partial x} \frac{\partial^2 u}{\partial x^2} dx \\ & - \int_{\Omega} \left[\left(\frac{\partial u_1}{\partial x}(0, t) \right)^{\frac{1}{2}} - \left(\frac{\partial u_2}{\partial x}(0, t) \right)^{\frac{1}{2}} \right] G' \frac{\partial u_2}{\partial x} \frac{\partial^2 u}{\partial x^2} dx \\ & - \int_{\Omega} (v_T(u_1) - v_T(u_2)) \frac{\partial^2 u_2}{\partial x^2} \frac{\partial u}{\partial x} dx. \end{aligned} \right. \quad (2.14)$$

We remark that

$$\left| \left(\frac{\partial u_1}{\partial x}(0, t) \right)^{\frac{1}{2}} - \left(\frac{\partial u_2}{\partial x}(0, t) \right)^{\frac{1}{2}} \right| \leq \frac{\left| \frac{\partial u}{\partial x}(0, t) \right|}{\alpha}$$

(α given by Corollary 2.2) and

$$\left| \frac{\partial u}{\partial x}(0, t) \right| \leq \sqrt{2} \left\| \frac{\partial u}{\partial x}(t) \right\|^{\frac{1}{2}} \left\| \frac{\partial^2 u}{\partial x^2}(t) \right\|^{\frac{1}{2}}.$$

Plugging this estimate in the R.H.S. of (2.14), it comes out, after several computations based on Young's formula, a relation of the form:

$$\frac{d}{dt} \left\| \frac{\partial u}{\partial x}(t) \right\|^2 \leq M(t) \left\| \frac{\partial u}{\partial x}(t) \right\|^2 \quad (2.15)$$

where $M \in L^1(0, S)$. By Gronwall's lemma, $\left\| \frac{\partial u}{\partial x}(t) \right\| \equiv 0$, hence

$$\frac{\partial u_1}{\partial x} = \frac{\partial u_2}{\partial x}$$

in Q and the uniqueness. ■

Acknowledgements: The authors are indebted to Professor M. Crandall who brought J. Moser's results to their attention.

FOOTNOTES

- (1) See C. Lainé [23].
 - (2) A weighting process is introduced with the contraction $k_\ell \phi_{\ell j i}$ which minimizes the role played by the large anisotropic structures.
 - (3) Recall that $H^m(\Omega)$ is the subset of functions of $L^2(\Omega)$ whose partial derivatives up to the order m are also in $L^2(\Omega)$. $H^1_0(\Omega)$ is the closure of the set of smooth functions $D(\Omega)$ in $H^1(\Omega)$, and $H^{-1}(\Omega)$ its dual space. If X is a Banach space, $L^p(0, S; X)$ is the set of functions of $]0, S[$ into X of power p integrable, $1 \leq p < +\infty$ (essentially bounded if $p = +\infty$).
 - (4) V_m = finite dimensional space spanned by w_1, \dots, w_m .
-

REFERENCES

- [1] Glowinski, R., Mantel, B., Periaux, J., and Pironneau, O., " H^{-1} least squares method for the Navier-Stokes equations," 1st Int. Conf. on Numerical Methods in Laminar and Turbulent Flow, Eds. Taylor, C., Morgan, K. and Brebbia, C. A. (Pentech Press, Swansea, 1978).
- [2] Bradshaw, P., Ferris, D. H., and Atwell, N. P., "Calculation of boundary development using the turbulent energy equation," J.F.M. 28 (1967) p. 593-616.

- [3] Launder, B. E., and Spalding, D. B., Lecture in "Mathematical Models of Turbulence," (Academic Press, New-York, 1972).
- [4] Jeandel, D., Brison, J. F., and Mathieu, J., "Modeling methods in physical and spectral space," Physics of Fluids, December 1977.
- [5] Hopf, E., J. Ratl. Mech. Anal. (1952), p. 1-87.
- [6] Monin, A. S., and Yaglom, A. M., "Statistical fluid mechanics: mechanics of turbulence," The M.I.T. Press, Vol. 2, 1967.
- [7] Lundgreen, T. S., Physics of Fluids 10, (1967), p. 969.
- [8] Dopazo, _., and O'Brien, E. E., Combusion Science and Technology, Vol. 19 (1975) p. 99-122.
- [9] Mathieu, J., "Idées actuelles sur la turbulence," CANCAM 79, Sherbrooke (1979), p. 465-502.
- [10] Lumley, J., "Von Karman Institute," J. 1975, Lectures series 76.
- [11] Launder, B. E., Reece, G. J., and Rodi, W., "Progress in the development of a Reynolds Stress closure," J. Fluid Mech. 68, p. 537-566.
- [12] Cea, J., and Geymonat, G., "Une méthode de linéarisation via l'optimisation," Instituto Nazionale di Alta Matematica, Symposia Mathematica, Vol. 10 (1972).
- [13] Gence, J. N., and Mathieu, J., "On the application of successive plane strains to guid generated turbulence," J. Fluid Mech. (1979), Vol. 93, Part 3, p. 501-513.
- [14] Proudman, I., and Reid, W. H., "On the decay of a normally distributed and homogeneous turbulent velocity field," Phil. Trans. Roy. Soc. A, (1954), p. 163-189.
- [15] Orszag, S. A., "Analytical theories of turbulence," J. Fluid Mech. 41, (1970), p. 262-386.
- [16] Cambon, C., Jeandel, D., and Mathieu, J., "Spectral modeling of homogeneous non isotropic turbulence," Journal of Fluid Mechanics, Vol. 104 (1981), p. 247-262.
- [17] Deardorff, J. W., "A numerical study of three-dimensional turbulent channel flow at large Reynolds numbers," J. Fluid Mech. 41 (1970), p. 453.
- [18] Leonard, A., "Energy cascade in large-eddy simulations of turbulent fluid flows," Adv. in Geophys. A 18 (1974), p. 237.
- [19] Smagorinsky, J. S., "General circulation experiments with the primitive equations. I: the basic experiment," Mon. Weath. Rev. 91 (1963) p. 99.
- [20] Gence, J. N., and Matheiu, J., "The return to isotropy of an homogeneous turbulence having been submitted to two successive plane strains," J.F.M., Vol. 101, Part 3 (1980), p. 555-556.
- [21] Ferziger, J. H., "Higher-level simulations of turbulent flows," V.K.I. lecture series 5 (1981).

- [22] Bennani, A., Alcaraz, E., and Mathieu, J., "An experimental setup for interaction turbulence-chemical reaction research," To be published in Int. J. of Heat and Mass Transfer.
- [23] Lainé, C., Etude mathématique et numérique de modèles de turbulence en conduite plane, Thèse, Lyon (1980).
- [24] Lainé, C., C. R. Acad. Sc. Paris, Série B, 291 (1980), p. 63-66.
- [25] Brauner, C. M. and Lainé, C., "Mathematical and numerical study of a turbulent viscosity model," to appear.
- [26] Adams, J. E., Sobolev spaces, Academic Press, New York (1975).
- [27] Lions, J. L., "Quelques méthodes de résolution de problèmes aux limites non linéaires," Dunod, Paris (1968).
- [28] Lions, J. L. and Magenes, E., "Problèmes aux limites non homogènes et applications," Dunod, Paris (1968).
- [29] Moser, J., "A Harnack inequality for parabolic differential equations," Comm. Pure Appl. Math. 17 (1964), p. 101-134, and correction in 20 (1967), p. 231-236.
- [30] Moser, J., "On a pointwise estimate for parabolic differential equations," Comm. Pure Appl. Math. 24 (1971), p. 727-740.
- [31] Protter, M. H. and Weinberger, H. F., "Maximum principle in differential equations, Prentice Hall (1967).

INSTABILITY OF PIPE FLOW

Anthony T. Patera and Steven A. Orszag

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Mass. 02139

The stability of axisymmetric disturbances in Hagen-Poiseuille (pipe) flow to infinitesimal non-axisymmetric perturbations is studied by numerical solution of the Navier-Stokes equations. It is shown that, if the axisymmetric flow is allowed to evolve according to the Navier-Stokes equations, the non-axisymmetric instability is strong for Reynolds numbers $R \gtrsim 3000$ but weak for $R \lesssim 1000$, the cut-off being due both to increased decay of the axisymmetric flow and reduced growth of the non-axisymmetric perturbation. The non-axisymmetric instability can be isolated from the very fast axisymmetric decay if the axisymmetric flow is forced in such a way that its energy is maintained constant, but phase relations are allowed to develop. In this case, it is found that the Reynolds number at which transition occurs in experiments, $R \approx 2000$, delimits the regions of inviscid (i.e. R -independent) growth and viscosity-dominated decay.

INTRODUCTION

Experiments indicate that, unless great care is taken to remove background disturbances, Hagen-Poiseuille (pipe) flow undergoes transition to turbulence at Reynolds numbers R (based on pipe radius and centerline velocity) of the order of 2000 (Wynanski & Champagne 1973). Experimental work has successfully addressed the problem of relating transition structures (e.g. puffs, slugs) to fully-developed turbulent flow (Rubin et al. 1979), but there has been little progress experimentally or analytically towards finding and characterizing an instability which is significant at the low Reynolds numbers at which transition occurs naturally.

It is now generally accepted that Hagen-Poiseuille flow is stable to all linear axisymmetric and non-axisymmetric disturbances (Davey & Drazin 1969; Metcalfe & Orszag 1973; Salwen et al. 1980). Thus nonlinear interactions must be involved if experimental observations are to be explained. Most previous finite-amplitude stability work has been axisymmetric in nature. The amplitude expansion techniques of Stuart (1971) [extended to problems where no neutral curve exists (Reynolds & Potter (1967))] have been used to suggest the existence of axisymmetric equilibria and, hence, axisymmetric subcritical instability (Davey & Nguyen 1971; Itoh 1977). However, Patera & Orszag (1981) have shown numerically that these solutions do not exist. It was found that, in the region of wavenumber (α), Reynolds number (R) space in which the search was conducted, pipe flow is stable to all finite-amplitude axisymmetric disturbances. We suspect that pipe flow is globally stable to all linear and nonlinear axisymmetric disturbances at all (α, R) .

Thus not only finite-amplitude but also non-axisymmetric disturbances must be considered before an instability can be obtained in Hagen-Poiseuille flow. We

discuss this situation briefly in the context of plane channel flows. In plane Poiseuille flow, two-dimensional equilibria do exist (Zahn et. al. 1974; Herbert 1976); by themselves, they can not explain transition inasmuch as they only exist for $R \geq 2900$ (while transition may occur at $R = 1000$) and they result in saturated states, not chaotic motion. However, we have previously shown (Orszag & Patera 1980, 1981), that the two dimensional states are strongly unstable to three-dimensional perturbations. Furthermore, very slowly-decaying two-dimensional states (quasi-equilibria) exist down to $R \approx 1000$. To the extent that three-dimensional finite-amplitude flows are typically chaotic in shear flows (Orszag & Kells 1980), this isolates the basic mechanism of transition in terms of finite-amplitude two-dimensional disturbances and infinitesimal three-dimensional perturbations. In contrast to plane Poiseuille flow, in plane Couette flow there are neither equilibria nor quasi-equilibria. However, it was found (Orszag & Patera 1981) that the same mechanism that operates in plane Poiseuille flow also holds in plane Couette flow, i.e. two-dimensional decaying finite-amplitude states are strongly unstable to three-dimensional infinitesimal perturbations.

The three-dimensional instability is not yet completely understood, although it is clear from numerical experiments that it is inviscid in nature, in contrast to the two-dimensional equilibria (and their linear counterpart, Tollmien-Schlichting waves), which require finite viscosity. Note, however, that this inviscid three-dimensional instability is not a classical inflectional instability. For, if the instantaneously inflectional two-dimensional profiles (generated by the equilibria) drive the instability, then one would also expect two-dimensional perturbations to the equilibria to be unstable, which is not the case.

The success of the two-dimensional finite-amplitude/three-dimensional infinitesimal structure in explaining subcritical transition in plane channel flows suggests that such a mechanism may also prove relevant in pipe Poiseuille flow, especially since pipe Poiseuille flow and plane Couette flow are very similar in terms of both linear theory and two-dimensional finite-amplitude behavior. In this paper, we show that this is indeed the case. Furthermore, new insights are provided into subcritical linear three-dimensional instability, and new tools are presented for carrying out the stability analysis when two-dimensional equilibria do not exist.

THREE-DIMENSIONAL LINEAR PROBLEM

The problem to be studied here involves the stability of Hagen-Poiseuille flow (defined as flow in a circular pipe of radius 1 driven by the constant pressure gradient $-4/R$) to finite-amplitude axisymmetric and infinitesimal three-dimensional (non-axisymmetric) perturbations:

$$\vec{v}(x, r, \theta, t) = \vec{v}^{(0)}(x, r, t) + \epsilon \vec{v}^{(1)}(x, r, \theta, t). \quad (1)$$

Here x, r , and θ are the axial, radial, and azimuthal coordinates, respectively, and ϵ is a (small) parameter related to the amplitude of the three-dimensional component. With the decomposition (1) (and $\epsilon \ll 1$), the Navier-Stokes equations are separable in θ so that it is consistent to assume that the θ -dependence of $\vec{v}^{(1)}$ is of the form $e^{\pm i m \theta}$. If \vec{v} is also expanded in a Fourier series in x (keeping all modes in this direction because of nonlinear interactions), the velocity is represented as

$$\vec{v}(x, r, \theta, t) = \sum_{n=-\infty}^{\infty} \sum_{q=-1}^1 \epsilon |q| \vec{v}_n^{(q)}(r, t) e^{i \alpha n x} e^{i q m \theta} \quad (2a)$$

where $2\pi/\alpha$ is the periodicity interval in the x -direction. Reality of the

velocity field implies the following symmetry:

$$\vec{v}_{-n}^{(-q)} = \vec{v}_n^{(q)*} \quad (2b)$$

where * denotes complex conjugate. Note the assumption of periodicity in x and θ corresponds in classical stability theory to a temporal stability analysis.

The incompressible Navier-Stokes equations govern the evolution of \vec{v} in time. To $O(\epsilon^2)$, the equations for $\vec{v}_n^{(q)}$ are

$$\frac{\partial \vec{v}_n^{(q)}}{\partial t} = \sum_{\ell=-\infty}^{\infty} \{ \vec{v}_\ell^{(0)} \times \vec{\omega}_{n-\ell}^{(q)} + (1-\delta_{0q}) \vec{v}_\ell^{(q)} \times \vec{\omega}_{n-\ell}^{(0)} \} \quad (3a)$$

$$-\vec{v}_n^{(q)} + \frac{4}{R} \delta_{0q} \delta_{0n} \hat{x} + \frac{1}{R} \nabla^2 \vec{v}_n^{(q)}$$

$$\vec{\nabla} \cdot \vec{v}_n^{(q)} = 0 \quad (3b)$$

$$\vec{\omega}_n^{(q)} = \vec{\nabla} \times \vec{v}_n^{(q)} \quad (3c)$$

where δ_{ij} is the Kronecker delta function and the carat denotes a unit vector. The boundary conditions on \vec{v} are

$$\vec{v}_n^{(q)} (r=0) \text{ bounded}$$

$$\vec{v}_n^{(q)} (r=1) = 0,$$

the latter being the no-slip condition. No boundary conditions are required in the other directions due to the assumption of periodicity.

Note that classical linear stability theory is a special case of the above formulation in which $\vec{v}^{(0)}(x,r,t) = (1-r^2)\hat{x}$. In this case, the relevant parameter space is (R, α, m) , and it is generally agreed that all disturbances decay. The solution to the linear characteristic-value problem will be denoted as $\{R, \alpha, m\}$. With nonlinear effects allowed, the parameter space must be extended to include a measure of amplitude, such as the modal energies:

$$E_n^{(q)} = 6 \int_0^1 |\vec{v}_n^{(q)} \cdot \vec{v}_n^{(q)}| r dr \quad (4a)$$

$$E_{\text{tot}}^{(q)} = \sum_{n=-\infty}^{\infty} E_n^{(q)} - \delta_{0q} E_0^{(0)} \quad (4b)$$

(The factor of 6 is so that energies for $n \neq 0$ are measured relative to that of the laminar parallel flow $(1-r^2)\hat{x}$). $E_{\text{tot}}^{(0)}$ and $E_{\text{tot}}^{(1)}$ represent the axisymmetric and non-axisymmetric perturbation energies, respectively.

NUMERICAL METHOD

The numerical technique used here for solving the Navier-Stokes equations is very similar to that discussed in detail in Patera & Orszag (1981a) for solving the finite-amplitude axisymmetric problem. The evaluation of the nonlinear terms, the time-stepping method, and the matrix inversion techniques all carry over directly to the non-axisymmetric problem. It is only in the treatment of the axis, and, in particular, its effect on the choice of suitable spectral expansions and boundary conditions in which special care must be taken in solving the three-dimensional problem. We limit discussion here to these more subtle points.

As in the axisymmetric problem, the radial dependence of \vec{v} is expanded in a Chebyshev polynomial series, where the p th Chebyshev polynomial is defined

$$T_p(\cos \theta) = \cos p\theta$$

The fact that the domain of interest is $0 \leq r \leq 1$ whereas the argument of the Chebyshev polynomials, z , satisfies $-1 \leq z \leq 1$ can be handled in one of two ways. First, a simple scaling map can be used to stretch the $[0,1]$ interval. Even better, a given variable can be expanded in an even or odd Chebyshev series, the parity of the expansion being determined by the behavior of the variable at the axis, $r = 0$. In particular, we require expansions for the axial, radial, and azimuthal velocities, denoted u, v , and w , respectively.

To determine the leading behavior of u, v , and w as $r \rightarrow 0$ we examine the most singular part of (3), namely the viscous terms. (The conditions derived in this way will also insure boundedness of the vorticity and divergence at $r = 0$). In cylindrical coordinates the viscous stress term is not diagonal, i.e. the stress in the x_k ($x_1 = x, x_2 = r, x_3 = \theta$) direction cannot be expressed solely in terms of v_k (k here again denoting directional component). However, a simple transformation of the independent variable $[u, v, w] \rightarrow [\tilde{u}, \tilde{v}, \tilde{w}]$ defined by

$$\begin{aligned}\tilde{u} &= u \\ \tilde{v} &= v + iw \\ \tilde{w} &= v - iw\end{aligned}\tag{5}$$

diagonalizes the viscous operator. For a velocity component of the form $\vec{v}(x, r, \theta) = \vec{v}_{nm} e^{i\alpha n x} e^{im\theta}$ the viscous term can now be expressed as

$$\nabla^2 \vec{v} = \begin{pmatrix} L - \frac{m^2}{r^2} & 0 & 0 \\ 0 & L - \frac{(m+1)^2}{r^2} & 0 \\ 0 & 0 & L - \frac{(m-1)^2}{r^2} \end{pmatrix} \begin{pmatrix} \tilde{u}_{nm} \\ \tilde{v}_{nm} \\ \tilde{w}_{nm} \end{pmatrix} e^{i\alpha n x} e^{im\theta}\tag{6a}$$

where

$$L = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) - \alpha^2 n^2\tag{6b}$$

Note the change of variable (5) is advantageous for the solution of the Navier-Stokes equations: in this form three separate Laplacians must be inverted, whereas using $[u, v, w]$ there is a coupling of two of the equations which would require (if the solution were fully implicit) additional computing time, storage, and complexity.

Analysis of local behavior near the singular point of (6) at $r = 0$ (restricting attention to $m \geq 0$ since reality relates m and $-m$) gives

$$\begin{aligned}\tilde{u} &\sim r^{\pm m}, & m > 0; & \quad 1, \ln r \quad m = 0 \\ \tilde{v} &\sim r^{\pm(m+1)}, & \text{all } m \geq 0 \\ \tilde{w} &\sim r^{\pm(m-1)}, & m \neq 1; \quad 1, \ln r \quad m = 1\end{aligned}\tag{7}$$

(Batchelor & Gill 1962; Orszag 1974).

Consistent with this behavior the complete spectral expansion of the velocity field can be written as

$$v_k(x, r, \theta, t) = \sum_{q=-1}^1 \sum_{n=-N}^N \sum_{p=0}^P \epsilon |q| v_{n,p,k}^{(q)} e^{i\alpha n x} e^{i q m \theta} T_\ell(r) \tag{8a}$$

where $\ell = 2p + b(k, m | q)$ and

$$b(r, s) = \begin{cases} 0 & (r = 1 \text{ and } s \text{ even}) \text{ or } (r = 2, 3 \text{ and } s \text{ odd}) \\ 1 & (r = 1 \text{ and } s \text{ odd}) \text{ or } (r = 2, 3 \text{ and } s \text{ even}) \end{cases} \tag{8b}$$

and k refers to directional component. Reality imposes the condition

$$v_{n,p,k}^{(q)} = v_{-n,p,k}^{(-q)*} \tag{8c}$$

on the finite Fourier series.

Finally, we discuss imposition of boundary conditions at $r = 0$. When the equation to be solved can be regularized, no boundary condition is required (i.e. the singularity in the equation is removed by rejecting the unbounded solution). If the equation cannot be regularized, a boundary condition at $r = 0$ is required. A boundary condition is easily obtained from (7) for either the solution or its derivative.

We will comment briefly here on the solution of the linear - theory characteristic value problem as linear modes provide the initial conditions for the finite

amplitude runs. As in the simulation, a pseudospectral Chebyshev method is used to set up the matrix equations, and either a cubically convergent (inverse iteration/Rayleigh quotient) local method or a global (QR) algorithm is used to solve the algebraic eigenvalue problem.

RESULTS

Before presenting finite-amplitude results, we comment that the numerical code accurately simulates linear behavior. Test runs indicate that at $R = 4000$, $\alpha = 1.0$, with $P = 32$, $\Delta t = 0.025$, the energy of both $m = 0$ modes, $\{R, \alpha, m\} = \{4000, 1.0, 0\}$, and $m = 1$ modes, $\{4000, 1.0, 1\}$, decay at a rate which agrees to within about 0.1% of linear theory. In addition to these linear tests, axisymmetric finite amplitude runs were also made to confirm the conclusions of our previous work on pipe flow (Patera & Orszag 1981).

In the finite-amplitude axisymmetric/linear non-axisymmetric runs, initial conditions are of the form

$$\vec{v}^{(0)}(x, r, t = 0) = \vec{v}_L(\{R, \alpha, 0\}) + (1 - r^2)\hat{x}$$

$$\vec{v}^{(1)}(x, r, \theta, t = 0) = \vec{v}_L(\{R, \alpha, 1\})$$

where the linear mode \vec{v}_L is chosen to be the least stable wall mode at the given $\{R, \alpha, m\}$. The initial energy of the axisymmetric disturbance is specified as $E_{\text{tot}}^{(0)} \equiv E_{\text{tot}}^{(0)}(t = 0)$. The initial amplitude of the non-axisymmetric component is of course irrelevant as the theory is linear in the azimuthal direction. Note that the results discussed below concern flows developing from initial conditions consisting of wall modes with $m = 1$. The behavior of center modes and modes with general $m (\neq 1)$ is not important and will be discussed briefly at the end of this section. For the range of R, α run here, the numerical parameters $P = 32$, $N = 8$, $\Delta t = 0.025$ provide fully converged results as was confirmed by doubling (or halving) the values given for P, N , and Δt above and comparing the resulting solutions.

The natural way to proceed in seeking a transition-related instability is to establish that an instability exists for Reynolds numbers slightly larger than R_T , the transitional Reynolds number, and following the behavior of the instability as R passes through R_T . In Fig. 1 we demonstrate that a three-dimensional instability of the type found previously in plane channel flows actually exists in Hagen-Poiseuille flow; the wavenumber dependence of the instability at $R = 4000$ is illustrated by a plot of $\ln E_{\text{tot}}^{(q)}$ for $q = 0, 1$ vs time. (Here $E_{\text{tot}}^{(0)} = 0.04$). For $\alpha = 0.5$ there is only weak coupling of the axisymmetric and non-axisymmetric fields and three-dimensional growth is less than for $\alpha = 1.0$ or $\alpha = 2.0$. The growth of non-axisymmetric disturbances at $\alpha = 2.0$ is approximately the same as that at $\alpha = 1.0$, but the axisymmetric solutions decay much more quickly. Thus, taking into consideration both axisymmetric decay and non-axisymmetric growth, $\alpha = 1.0$ appears to be nearly the most dangerous wavenumber.

In Fig. 2 we again plot $\ln E_{\text{tot}}^{(q)}$ for $q = 0, 1$ vs time but now at $\alpha = 1.0$ letting the Reynolds number vary from 500 to 4000. Although it is clear from this series of runs that the instability is dangerous at $R = 3000$ and insignificant at $R = 500$, a Reynolds number of 2000 (or slightly larger) is not singled in the plot as distinguishing growth from decay. This problem of inter-

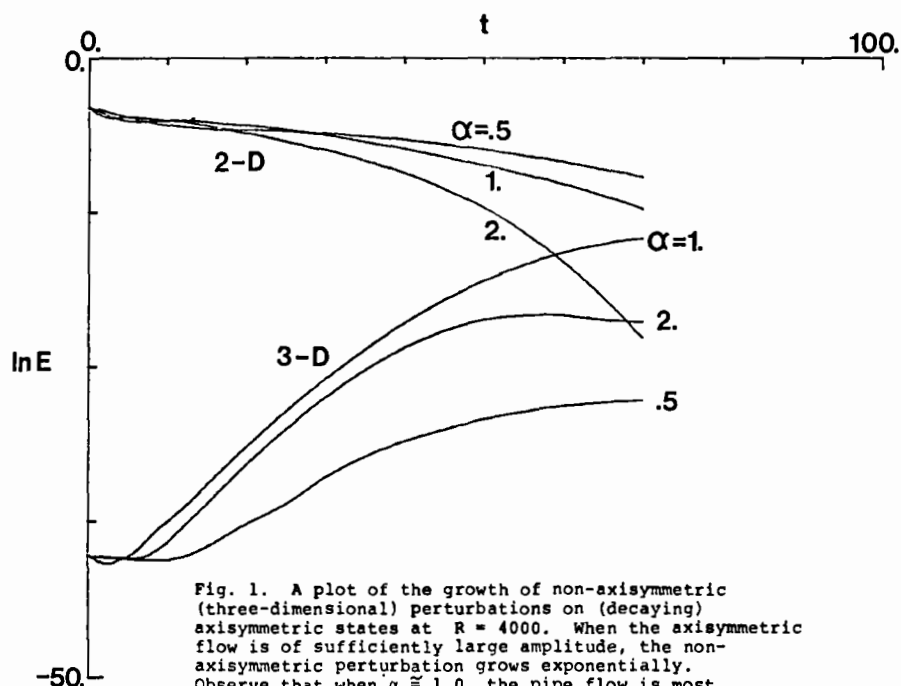


Fig. 1. A plot of the growth of non-axisymmetric (three-dimensional) perturbations on (decaying) axisymmetric states at $R = 4000$. When the axisymmetric flow is of sufficiently large amplitude, the non-axisymmetric perturbation grows exponentially. Observe that when $\alpha \approx 1.0$ the pipe flow is most unstable. Lower wavenumbers significantly decrease the growth of non-axisymmetric perturbations, while higher wavenumbers result in sharply increased axisymmetric decay.

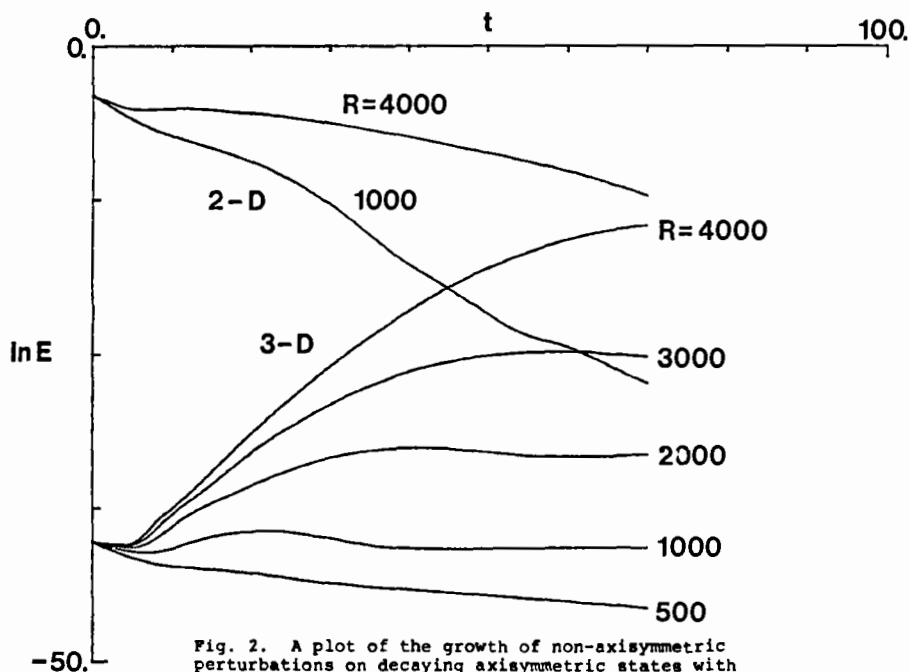


Fig. 2. A plot of the growth of non-axisymmetric perturbations on decaying axisymmetric states with $\alpha = 1.0$ for various R . The instability is strong for $R \geq 3000$ and weak for $R \leq 1000$, the cut-off due primarily to increased axisymmetric decay.

pretation arises from the fact that, in pipe flow, finite-amplitude axisymmetric disturbances decay on a time scale even shorter than those in plane Couette flow, and thus the details of the instability are obscured by the two-dimensional damping.

To isolate the three-dimensional instability requires forcing of the axisymmetric flow in such a way that it evolves naturally while maintaining a fixed amplitude. We term such a forcing of the two-dimensional flow A-frozen (amplitude -) frozen, meaning that overall amplitudes are constrained, but phase relations are not. To wit, the following scheme is used: after each time step, $E^{(0)}$ and $E_{tot}^{(0)}$ are computed. The $n=0$ mode is then scaled by $\sqrt{E^{(0)}/E_{tot}^{(0)}}$ and all $n \neq 0$ modes by $\sqrt{E_{tot}^{(0)}/E^{(0)}}$, where $E^{(0)}$ and $E_{tot}^{(0)}$ are the desired (constant) two-dimensional mean-flow and perturbation energies, respectively. Note that, with this scheme, harmonics can develop and waves can propagate. We are only imposing a weak energetic constraint on the axisymmetric flow.

In Fig. 3 we show a plot similar to Fig. 2 except here the axisymmetric field is A-frozen. Note first that the growth of three-dimensional modes is very closely pure exponential as would be expected if one actually had axisymmetric equilibria; thus we have effectively separated the effects of axisymmetric decay and non-axisymmetric growth. More importantly, note that in this plot $R = 2000$ delimits regions of inviscid (i.e. R -independent) growth ($R \geq 2000$) and viscous-dominated behaviour ($R \leq 2000$). This is reflected by the close spacing of the $\ln E_{tot}^{(1)}$ curves for $R \geq 2000$ and the abrupt change to wide spacing at $R \approx 2000$.

The fact that forcing is required to highlight the most significant aspects of the instability in Hagen-Poiseuille flow but not in plane Poiseuille flow is certainly linked to the different nature of transition in these two flows. The appearance of puffs and slugs in Hagen-Poiseuille flow experiments is evidence that explosively unstable transition structures may ultimately decay (barring very large disturbances), just as the behavior of our instability in the unforced case is numerical evidence of the same phenomenon.

Further insight into the present non-axisymmetric instability is obtained by investigating the requirements on the forcing protocol in order that three-dimensional instability be maintained. In particular, we define an axisymmetric field to be ϕ -frozen (phase -) frozen if, at every time step, the axisymmetric field is reset to its frozen value in the laboratory coordinate frame; note the difference between ϕ -frozen (where phase relations are not allowed to evolve naturally) and A-frozen. In Fig. 4, $\ln E_{tot}^{(1)}$ for A-frozen and ϕ -frozen fields are compared. Note that there is virtually no growth in ϕ -frozen case. Furthermore, it is not the lack of harmonic generation in the ϕ -frozen field that rules out growth, for, if the axisymmetric field is ϕ -frozen at some time T after it has evolved naturally for $0 \leq t < T$, three-dimensional growth rapidly turns off. The drastically different results obtained are no doubt due to the synchronization phenomenon found in plane Poiseuille flow (Patera & Orszag 1981b), where both the two-dimensional wave and the most dangerous three-dimensional perturbation travel at the same phase speed (i.e. the temporal eigenvalue σ is real). In the ϕ -frozen case this synchronization is inhibited.

We have not yet thoroughly investigated the instability of axisymmetric disturbances to higher azimuthal-wavenumber ($m > 1$) perturbations. However, it appears (as in linear theory) that the effect of increasing m is stabilizing. Also, runs have been made in which the initial conditions involve center modes rather than wall modes. It is found that, for the same initial maximum amplitude of the axisymmetric disturbance, wall modes are the more dangerous. (Note that if one compares non-axisymmetric growth due to a wall and center mode of the same initial energy, the latter appears more dangerous. However, this result is an artifact of the energy definitions (4) in which the energy of the

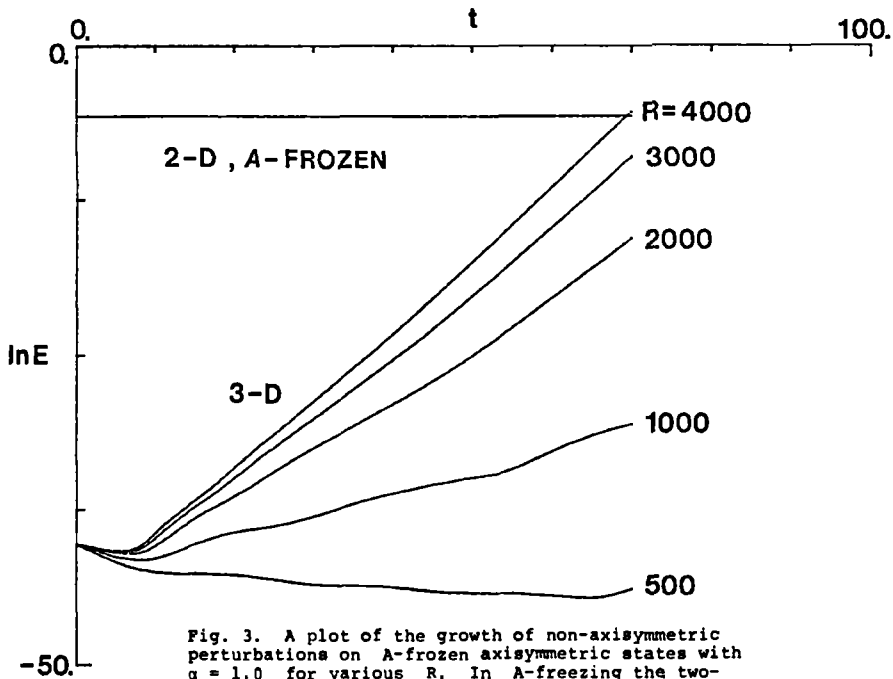


Fig. 3. A plot of the growth of non-axisymmetric perturbations on A-frozen axisymmetric states with $\alpha = 1.0$ for various R . In A-freezing the two-dimensional field we prevent axisymmetric decay but allow non-axisymmetric growth, thus isolating the mechanism of three-dimensional instability. It is seen that at $R \approx 2000$ there is a change from viscous behavior (strongly R -dependent) to inviscid growth (approximately R -independent).

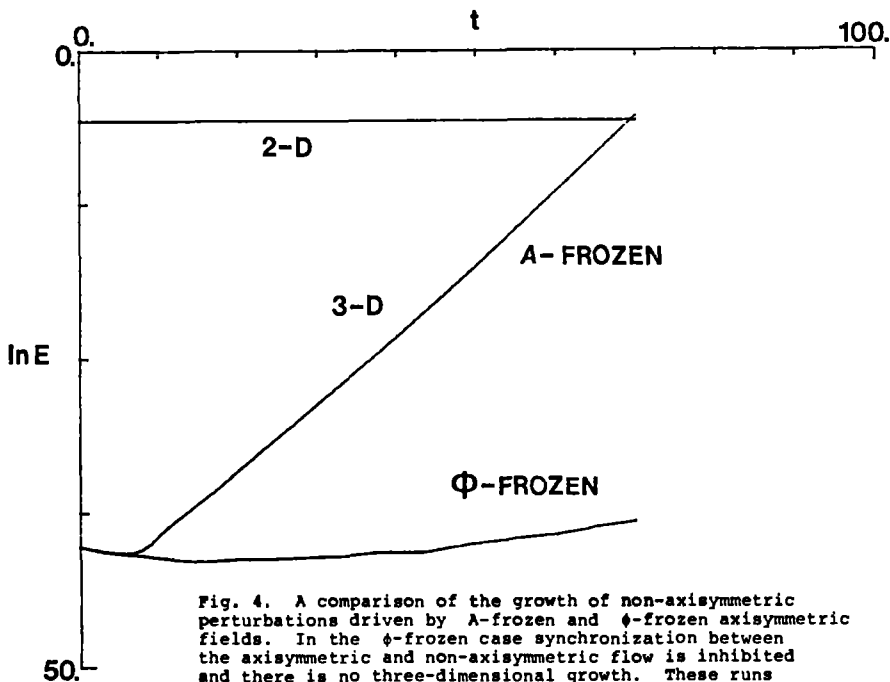


Fig. 4. A comparison of the growth of non-axisymmetric perturbations driven by A-frozen and Φ -frozen axisymmetric fields. In the Φ -frozen case synchronization between the axisymmetric and non-axisymmetric flow is inhibited and there is no three-dimensional growth. These runs are at $R = 4000$, $\alpha = 1.0$.

disturbance is weighted with distance from the axis. Thus very large amplitude center modes result in small energies. For example, an initial energy of 0.04 corresponds to roughly a 10% maximum amplitude for a wall mode, but to almost a 40% maximum amplitude for a center mode.)

This work was supported by the Office of Naval Research under Contracts No. N00014-77-C-0138 and No. N00014-79-C-0478. Development of the stability codes was supported by NASA Langley Research Center under Contract No. NAS1-16237.

REFERENCES:

- [1] Batchelor, F.K. & Gill, A. E. 1962 Analysis of the stability of axisymmetric jets. *J. Fluid Mech.* 14, 529.
- [2] Davey, A. & Drazin, P. G. 1969 The stability of Poiseuille flow in a pipe. *J. Fluid Mech.* 36, 209.
- [3] Davey, A. & Nguyen, H.P.F. 1971 Finite-amplitude stability of pipe flow. *J. Fluid Mech.* 45, 701.
- [4] Herbert, T. 1976 Periodic secondary motions in a plane channel. In *Proc. 5th Int. Conf. on Numerical Methods in Fluid Dynamics* (ed. A.I. van de Vooren and P. J. Zandbergen), p. 235, Springer.
- [5] Itoh, N. 1977 Nonlinear stability of parallel flows with sub-critical Reynolds number. Part 2. Stability of pipe Poiseuille flow to finite axisymmetric disturbances. *J. Fluid Mech.* 82, 469.
- [6] Metcalfe, R. W. & Orszag, S. A. 1973 Numerical calculations of the linear stability of pipe flows. Flow Research Report No. 25, Kent, Washington.
- [7] Orszag, S. A. 1974 Fourier series on spheres. *Mon. Wea. Rev.* 102, 56.
- [8] Orszag, S. A. & Kells, L. C. 1980 Transition to turbulence in plane Poiseuille and plane Couette flow. *J. Fluid Mech.* 96, 159.
- [9] Orszag, S. A. & Patera, A. T. 1980 Subcritical transition to turbulence in plane channel flows. *Phys. Rev. Letters*, 45, 989.
- [10] Orszag, S. A. & Patera, A. T. 1981 Three dimensional instability of plane channel flows. To be published.
- [11] Patera, A. T. & Orszag, S. A. 1981 Finite-amplitude stability of axisymmetric pipe flow. To appear in *J. Fluid Mech.*
- [12] Reynolds, W. C. & Potter, M. C. 1967 Finite-amplitude instability of parallel shear flows. *J. Fluid Mech.* 278 465.
- [13] Rubin, Y., Wignanski, I., & Haritonidis, J. H. 1979 Further observations on transition in a pipe. In *Laminar-Turbulent Transition*, IUTAM Symposium (ed. R. Eppler & H. Fasel), p. 17, Springer.
- [14] Salwen, H., Cotton, F. W., & Grosch, C. E. 1980 Linear stability of Poiseuille flow in a circular pipe, *J. Fluid Mech.*, 98, 273.

- [15] Stuart, J. T. 1971 Nonlinear stability theory, *Ann. Rev. Fluid Mech.* 3, 347.
- [16] Wygnanski, I. & Champagne, F. H. 1973 On transition in a pipe, Part 1. The origin of puffs and slugs and the flow in a turbulent slug. *J. Fluid Mech.* 59, 281.
- [17] Zahn, J.-P., Toomre, J., Spiegel, E. A. & Gough, D. O. 1974 Nonlinear cellular motions in Poiseuille channel flow, *J. Fluid Mech.* 64, 319.

This Page Intentionally Left Blank

SOME FORMALISM AND PREDICTIONS OF THE PERIOD-DOUBLING ONSET OF CHAOS

Mitchell J. Feigenbaum

Theoretical Division
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

We initially consider one-parameter families of maps f_λ with the property that for a convergent sequence of parameter values, the map possesses stable periodic orbits of successively increasing period:

$$\lambda_n \rightarrow \lambda_\infty ;$$

$$f_{\lambda_n} \text{ has a stable orbit of period } 2^n .$$

In an interval of parameter values around λ_∞ , a fixed-point theory provides universal limits for these maps when appropriately iterated and rescaled. The operator whose fixed point is considered is

$$Tf(x) = -\alpha f(f(-x/\alpha)) \quad (1)$$

where f has a quadratic extremum conventionally located at the origin. α is determined by requiring T to have a fixed point g :

$$Tg = g; \quad g(0) = 1, \quad g(x) = \psi(x^2), \quad \psi \text{ real analytic.}$$

About g , DT_g has a unique eigenvalue, δ , in excess of 1. Numerically,

$$\alpha = 2.5029..$$

$$\delta = 4.6692..$$

In terms of this fixed point g ,

$$\lim_{n \rightarrow \infty} (-\alpha)^n f_{\lambda_\infty}^{2^n}(x/(-\alpha)^n) = v g(x/v), \quad (2)$$

where v is an f -dependent magnification. We assume f has been suitably magnified so that we can take $v = 1$; the limit in (2) is in this sense universal. (Observe that the object on the left is $T^n f_{\lambda_\infty}$.)

Consider

$$x_n(\mu) = f_{\lambda_n(\mu)}^{2^n}(x_n(\mu)) \quad (3)$$

$$\mu = \frac{d}{dx} f_{\lambda_n(\mu)}^{2^n}(x_n(\mu)) \quad (4)$$

$\{\lambda_n(\mu)\}$ is a sequence of parameters for which f_λ has a period-doubling sequence of orbits all of stability μ . ($\mu = -1$ determines bifurcation values, $\mu = 0$ superstable values, etc.). For $|\mu| < 1$ $x_n(\mu)$ is an element of a stable 2^n -cycle, which we choose as that element of smallest modulus. For $|\mu| > 1$ $x_n(\mu)$ is the smallest element of an unstable cycle which is coexistent with an attractor of interest. For example, there is a value of $\mu < -1$ which is the slope of f^{2^n} at its smallest fixed point when there is a superstable 2^{n+1} cycle:

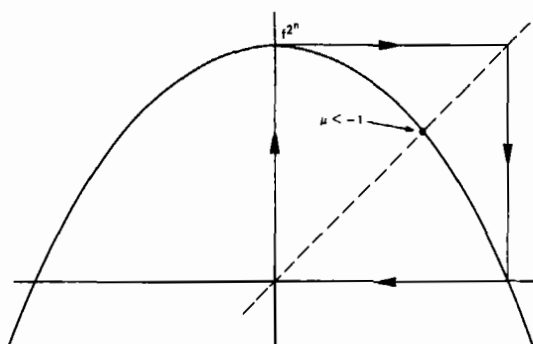


Figure 1: A schematic plot of f^{2^n} showing the point at which the slope equals $\mu < -1$.

For a more negative value of μ , we are considering an unstable fixed point of f^{2^n} when there is a superstable 2^{n+2} cycle. Similarly, there is a sequence of $\mu_r < -1$ such that a 2^{n+r} -cycle is superstable, until at $\hat{\mu}$ we consider the fixed point of $f_{\lambda_\infty}^{2^n}$ when period doubling has accumulated. Clearly,

$$x_n(\hat{\mu}) \sim \hat{x}/(-\alpha)^n$$

where

(5)

$$g(\hat{x}) = \hat{x}, \quad g'(\hat{x}) = \hat{\mu}.$$

Thus,

$$\hat{\mu} < \mu < -1$$

correspond to stable periodic behaviors below λ_∞ , while $\mu < \hat{\mu}$ for which f^{2^n} maps an interval about the origin onto itself:

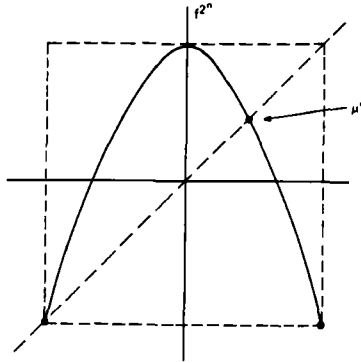


Figure 2: μ^* such that f^{2^n} maps an interval about the origin onto itself.

(Actually, there is a $\mu_n^* \rightarrow \mu^*$.) In this case $\lambda_n(\mu^*)$ will determine the "Misiurewicz-points" [1,2] at which ergodic and mixing behavior occurs. Accordingly μ serves as a "universal" parametrization of maps and the large n behavior of (3) and (4) shall allow a uniform presentation of both periodic and "chaotic" behaviors.

Returning to (3), consider

$$\begin{aligned} T^n f_{\lambda_{n+r}(\mu)} &= T^n (f_{\lambda_\infty} + (\lambda_{n+r}(\mu) - \lambda_\infty) \partial_\lambda f + \dots) \\ &= T^n f_{\lambda_\infty} + (\lambda_{n+r}(\mu) - \lambda_\infty) D T^n \cdot \partial_\lambda f + \dots \\ &\sim g + (\lambda_{n+r}(\mu) - \lambda_\infty) (D T_g)^n \cdot \partial_\lambda f + \dots \\ &\sim g + c(f) (\lambda_{n+r}(\mu) - \lambda_\infty) \cdot \delta^n \cdot \phi_\delta \end{aligned} \quad (6)$$

where

$$DT_{g \cdot \phi_\delta} = \delta \cdot \phi_\delta.$$

Since as $r \rightarrow \lim_{n \rightarrow \infty}$ is g , it follows that

$$\lim_{n \rightarrow \infty} c(f)(\lambda_{n+r}(\mu) - \lambda_\infty) \delta^n = \delta^{-r} k(\mu) \quad (7)$$

and

$$g_{r,\mu}(x) \equiv \lim_{n \rightarrow \infty} (-\alpha)^n f_{\lambda_{n+r}(\mu)}^{2^n}(x/(-\alpha)^n) \sim g + \delta^{-r} k(\mu) \cdot \phi_\delta. \quad (8)$$

$k(\mu)$ is independent of f since

$$T^r g_{r,\mu} = g_{0,\mu} \quad (9)$$

and $g_{0,\mu}$ must have slope μ at its smallest fixed point, which then determines $k(\mu)$. Equation (8) [3] is the strongest consequence of T 's fixed point, since it determines which limits of iterates exist and then that these limits are universal. For all applications, the strategy is to approximate iterates by these universal limits, where parameter values, amount of magnification, and order of iteration are "massaged" into the form of the left-hand side of (8). In this way Lyapunov exponents, trajectory scaling functions [4], etc. are determined. For example by (3),

$$(-\alpha)^n x_n(\mu) = (-\alpha)^n f_{\lambda_n(\mu)}^{2^n}((- \alpha)^n x_n(\mu)/(-\alpha)^n) = g_{0,\mu}((- \alpha)^n x_n(\mu)) \text{ by (8).}$$

Accordingly, denoting $\hat{x}(\mu)$ as the smallest fixed point of $g_{0,\mu}$:

$$\hat{x}(\mu) = g_{0,\mu}(\hat{x}(\mu)). \quad (10)$$

Then we have,

$$x_n(\mu) \approx \hat{x}(\mu)/(-\alpha)^n \quad (11)$$

demonstrating that α scaling applies whatever the stability. (It immediately follows from (4) that

$$\mu = g_{0,\mu}'(\hat{x}(\mu)) \quad (12)$$

which, as mentioned above, is the fact used to determine $k(\mu)$. Also, by (7),

$$c(f)(\lambda_n(\mu) - \lambda_\infty) \sim \delta^{-n} k(\mu) \quad (13)$$

which, of course, demonstrates that the parameter convergence rate δ is independent of stability. (In particular, the Misiurewicz values also accumulate to λ_∞ at the rate δ .)

It remains to determine $k(\mu)$. For μ sufficiently near $\hat{\mu}$ (8) can be used even for $r = 0$. For a larger range of μ (8) is used for $r = 1$ and then (9) is used to reach $r = 0$. (A precise computation of $g_{0,\mu}$ is ultimately deeply nonlinear so that numerics must be employed if approximations are unacceptable.) We demonstrate the simplest case:

$$g_{0,\mu} \sim g + k(\mu)\phi_\delta. \quad (14)$$

By (12)

$$\mu \approx g'(\hat{x}(\mu)) + k(\mu)\phi'_\delta(\hat{x}(\mu)) \approx \hat{\mu} + (\hat{x}(\mu) - \hat{x})g''(\hat{x}) + k(\mu)\phi'_\delta(\hat{x})$$

i.e.,

$$k(\mu)\phi'_\delta(\hat{x}) \approx (\mu - \hat{\mu})(1 - \hat{x}'(\hat{\mu})g''(\hat{x})) . \quad (15)$$

Next, by (10)

$$\begin{aligned} \hat{x}(\mu) &\cong \hat{x} + (\mu - \hat{\mu})\hat{x}'(\hat{\mu}) \approx g(\hat{x} + (\mu - \hat{\mu})\hat{x}'(\hat{\mu})) + k(\mu)\phi_\delta(\hat{x}) \\ &= \hat{x} + (\mu - \hat{\mu})\hat{\mu}\hat{x}'(\hat{\mu}) + k(\mu)\phi_\delta(\hat{x}) \end{aligned}$$

or

$$(\mu - \hat{\mu})(1 - \hat{\mu})\hat{x}'(\hat{\mu}) \approx k(\mu)\phi_\delta(\hat{x}).$$

Combined with (15),

$$k(\mu) \approx (\mu - \hat{\mu}) \left\{ \phi'_\delta(\hat{x}) + \frac{g''(\hat{x})\phi_\delta(\hat{x})}{1 - \hat{\mu}} \right\} . \quad (16)$$

Since g and ϕ_δ are determined (by solution of the fixed point equations), everything in (16) is universal. (Together with (13), it is also clear that if period-doubling sequences of periodic behavior converge to λ_∞ from below, then for all $\mu < \hat{\mu}$, sequences must converge from above.)

As an example of the utility of (8), we compute the Lyapunov exponent in the vicinity of λ_∞ .

$$\ell \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \ln \left| \frac{d}{dx} f^n \right|$$

At $\lambda_n(\mu)$, according to (8) f^{2^n} can be approximated by $g_{0,\mu}$. So, consider

$$\ell_n = \lim_{p \rightarrow \infty} \frac{1}{2^{n+p}} \ln \left| \frac{d}{dx} f_{\lambda_n}^{2^{n+p}}(x_0) \right|. \quad (17)$$

Define

$$x_{r+1} \equiv f_{\lambda_n}^{2^n}(x_r). \quad (18)$$

Then, by the chain rule,

$$\frac{d}{dx} f_{\lambda_n}^{2^{n+p}}(x_0) = \prod_{r=0}^{2^p-1} \frac{d}{dx} f_{\lambda_n}^{2^n}(x_r)$$

and

$$\ell_n = \frac{1}{2^n} \left| \lim_{p \rightarrow \infty} \frac{1}{2^p} \sum_{r=0}^{2^p-1} \ln \left| \frac{d}{dx} f_{\lambda_n}^{2^n}(x_r) \right| \right| \equiv \frac{1}{2^n} L(\mu). \quad (19)$$

Next, taking a cue from (8), define

$$\xi_r = (-\alpha)^n x_r.$$

By (12) and (8), $\xi_{r+1} \approx g_{1,\mu}(\xi_r)$ for large enough n .

Differentiating (8),

$$\frac{d}{dx} f_{\lambda_n}^{2^n}(x_r) \approx g'_{0,\mu}(\xi_r).$$

Accordingly, by (19),

$$L(\mu) \approx \lim_{p \rightarrow \infty} \frac{1}{2^p} \sum_{r=0}^{2^p-1} \ln |g'_{o,\mu}(\xi_r)|. \quad (20)$$

Thus, $L(\mu)$ is universal, while for large n , by (19), ℓ_n scales simply by powers of 2. For stable periodic behavior ($\mu > \hat{\mu}$), $\xi_r = \hat{x}(\mu)$ while by (12)

$$L(\mu) = \ln|\mu| < 0.$$

Finally, by (13) if $k(\mu)$ is available, $L(\mu)$ can be obtained as a function of $\lambda - \lambda_\infty$. Thus, $L(\mu)$ is a universal function periodic in the variable $\log_\delta(\lambda - \lambda_\infty)$ with period 1 for $\lambda < \lambda_\infty$. For $\mu = \mu^*$, $L(\mu^*)$ is evaluated from g_{o,μ^*} and the invariant density at μ^* from (20) and the ergodic theorem. $\ln(\mu^*)$ achieves the values $2^{-n}L(\mu^*)$ at $\lambda_n(\mu^*)$, so that for $\lambda > \lambda_\infty$ ℓ lies within a curve [5]

$$\ell = L(\mu^*) \cdot (\delta^{-n})^{\frac{\ln 2}{\ln \delta}}$$

or

$$\ell \sim L(\mu^*) \left| \frac{c(f)}{k(\mu^*)} (\lambda - \lambda_\infty) \right|^{\frac{\ln 2}{\ln \delta}}$$

At this point we enlarge our scope from maps on an interval to general multi-dimensional, dissipative dynamical systems. For a wide class of such systems it turns out that the mapping theory discussed above carries over without modifications (see reference [9] for details and further general information). However, beyond the usual 4.669 convergence rate, the dynamical consequences of the theory are most naturally phrased in the format of a scaling theory. This scaling formalism, while constructed from an "underlying" map, has the strong advantage of being transparent to any mention of a map. We proceed now in a more "physical" vein to exposit this formalism and determine its experimentally verifiable consequences.

Simply put, as a control parameter is varied, any dynamical variable demonstrates a systematic doubling of its period prior to the aperiodic limit signalling the onset of chaos.

However, since the period doublings accumulate so quickly, gross features remain nearly constant, while the new temporal features characterizing the newly

doubled period contain a small fraction of the energy. Moreover, as period-doubling is periodic in terms of logarithmic deviations from the transition point, each new modification is a constant fraction of its predecessor. This entails very definite consequences, whether viewed temporally or through a Fourier transform, which we now explore.

The basic theoretical quantification of this notion is that the deviations from the old periodicity that determine the new and doubled period systematically (and universally) scale from one doubling to the next. Formally, write

$$d_n(t) = \frac{x_n(t) - x_n(t+T_{n-1})}{2} \quad (21)$$

where $x_n(t)$ is any coordinate when the parameter is such that the n -th period doubling has occurred, for which the period is

$$T_n = 2 \cdot T_{n-1} . \quad (22)$$

(Actually, as the parameter varies T_n is only approximately constant. However, the parameter values accumulate, (22) is asymptotically correct.) Equation (21) shows d_n to be nonzero if the period of x_n is really T_n and not T_{n-1} , and so d_n is the deviation in the trajectory associated with the doubling period. The theoretical result (which is published elsewhere [4]) is that for large n ,

$$d_{n+1}(t) \sim \sigma(t/T_{n+1})d_n(t) . \quad (23)$$

so that σ has the symmetry

$$\begin{aligned} \sigma(x + \tfrac{1}{2}) &= -\sigma(x) \\ \sigma(x + 1) &= \sigma(x) . \end{aligned} \quad (24)$$

Equation (23) is a very strong prediction: acquire the temporal data $x_n(t)$ and compute from it the half-period deviation function of (21); similarly obtain d_{n+1} and divide it by d_n ; then for each n on an appropriately scaled plot, the same function $\sigma(x)$ should be observed. This is already a strong scaling result. However, there is an even stronger feature being tested: namely that $\sigma(x)$ is a known, available function with the approximate value

$$\sigma(x) = \begin{cases} \alpha^{-2} = 1/6.25\dots & 0 < x < .25 \\ \alpha^{-1} = 1/2.5029\dots & .25 < x < .5 \end{cases} \quad (25)$$

All spectral tests for period-doubling universality are tests of (23) while, at present, the results following from the approximation (25) are as precise as experiments can hope to resolve.

There are, however, difficulties in employing (23) as an experimental test, which follow from the requirement of making two sets of measurements at different parameter values. The simpler problem is the phasing of the two time bases: to use (23) the time origins must be the same. This is most easily handled by varying a delay, in say $d_n(t)$, until its zero crossings are most nearly coincident with those of $d_{n+1}(t)$. (There are indeed many nearly coincident crossings. Also, the time scale of $d_n(t)$ might require some adjustment so that d_{n+1} has precisely twice the period of d_n .) More seriously, the parameter values must be chosen so that the limit cycles of x_n and x_{n+1} have identical stability. This requires, for example, some convergence data which might be hard to come by unless successive parameter values at the bifurcation points (when the period doubles) are used. If this difficulties are attended to, however, (23) and (25) then serve as a simple and very direct measurement of the universality theory. (These same difficulties afflict the Fourier measurements, except that phasing is obviously unimportant if only amplitude spectra are determined.)

In fact, the universal function σ is different for different stabilities (e.g., at superstable values as opposed to bifurcation values of the parameter). Indeed, for each stability μ , the appropriate σ is computed from the functions $g_{r,\mu}$ of (8). However, the differences show up only in the corrections to the approximation (25). Indeed there are other functions σ agreeing to the level of (25) which determine deviation signals, $d(t)$, pertaining to the same parameter value (typically to be used at the transition value) which then bypass the above difficulties. We shall explore these tests after determining Fourier properties which shall motivate them.

Corresponding to the deviation signal, the fundamental frequency of x_n is $1/T_n = \frac{1}{2}(1/T_{n-1})$, that is, a half subharmonic of the previous fundamental. Had $d_n(t)$ vanished, there would have been no components at the odd multiples of the new fundamental. The even multiples represent the components at the previous fundamental; since the parameter change becomes increasingly small, these components cannot suffer significant changes. Accordingly, the basic result of the doubled period is the set of spectral lines at the odd multiples of the new subharmonic fundamental. These are determined not by all of $x_n(t)$, but only by

its "subharmonic part" $d_n(t)$. Since these scale (by σ), it is then clear that these successive additions to the spectrum geometrically decrease in a universal way, as determined by the Fourier analysis of σ . Let us now make these observations quantitative.

By definition,

$$\hat{x}_{(n)}(p) = \frac{1}{T_n} \int_0^{T_n} dt x_n(t) e^{2\pi i p t / T_n} \quad (26)$$

with inverse,

$$x_n(t) = \sum_p \hat{x}_{(n)}(p) e^{-2\pi i p t / T_n} . \quad (27)$$

Manipulating (26)

$$\hat{x}_{(n)}(2p+1) = \frac{1}{T_{n-1}} \int_0^{T_{n-1}} dt \frac{x_n(t) - x_n(t + T_{n-1})}{2} e^{2\pi i \frac{(2p+1)}{T_n} t} \quad (28)$$

or

$$\hat{x}_{(n)}(2p+1) = \frac{1}{T_{n-1}} \int_0^{T_{n-1}} dt d_n(t) e^{2\pi i \frac{2p+1}{T_n} t} \quad (29)$$

with inverse

$$d_n(t) = \sum_p \hat{x}_{(n)}(2p+1) e^{-2\pi i \frac{2p+1}{T_n} t} . \quad (30)$$

Thus, the spectral components at odd multiples of the subharmonic fundamental are determined solely by the deviation signal $d_n(t)$.

Since all the new subharmonic components arise from the same time signal, they can be smoothly connected as an ensemble by interpolating them. The natural interpolation that reflects their common source is, then, the analytic continuation provided by (29):

(31)

$$\hat{x}_n(\omega) = \frac{1}{T_{n-1}} \int_0^{T_{n-1}} dt d_n(t) e^{2\pi i \omega t}$$

$$\text{with} \quad \hat{x}_n\left(\frac{2p+1}{T_n}\right) = \hat{x}_{(n)}(2p+1)$$

Utilizing (30), this interpolation is

$$\hat{x}_n(\omega) = \sum_p \hat{x}_{(n)}(2p+1) \frac{1}{T_{n-1}} \int_0^{T_{n-1}} dt e^{-2\pi i (2p+1 - \omega T_n) t / T_n} \quad (32)$$

or

$$\hat{x}_n(\omega) = (1 + e^{\pi i \omega T_n}) \frac{1}{\pi i} \sum_p \frac{\hat{x}_{(n)}(2p+1)}{2p+1 - \omega T_n} \quad (33)$$

We now ask how $\hat{x}_{n+1}(\omega)$ is related to $\hat{x}_n(\omega)$. By definition

$$\hat{x}_{n+1}(\omega) = \frac{1}{T_n} \int_0^{T_n} dt d_{n+1}(t) e^{2\pi i \omega t}$$

Employing the scaling formula for large n , we have

$$\begin{aligned} \hat{x}_{n+1}(\omega) &\sim \frac{1}{T_n} \int_0^{T_n} dt \sigma\left(\frac{t}{2T_n}\right) d_n(t) e^{2\pi i \omega t} \\ &= \sum_p \hat{x}_{(n)}(2p+1) \frac{1}{T_n} \int_0^{T_n} dt \sigma\left(\frac{t}{2T_n}\right) e^{-2\pi i (2p+1 - \omega T_n) \frac{t}{T_n}} \\ &= \sum_p \hat{x}_{(n)}(2p+1) \frac{1}{2T_{n-1}} \int_0^{T_{n-1}} dt \left[\sigma\left(\frac{t}{2T_n}\right) - e^{\pi i \omega t} \sigma\left(\frac{t}{2T_n} + \frac{1}{4}\right) \right] e^{-2\pi i (2p+1 - \omega T_n) t / T_n} \end{aligned}$$

If we now use the approximation of (25), then

$$\hat{x}_{n+1}(\omega) \approx \frac{1}{2\alpha} \left(\frac{1}{\alpha} - e^{\pi i \omega T_n} \right) \sum_p \hat{x}_{(n)}(2p+1) \frac{1}{T_{n-1}} \int_0^{T_{n-1}} dt e^{-2\pi i (2p+1 - \omega T_n) t / T_n}$$

$$\hat{x}_{n+1}(\omega) \approx \frac{1}{2\alpha} \left(\frac{1}{\alpha} - e^{\pi i \omega T_n} \right) \hat{x}_n(\omega) \quad (34)$$

by use of the interpolation formula (32).

Equation (34) is the basic result of this paper regarding Fourier analysis, and is rich in content [6]. The first conclusion to be drawn is that rather than scaling uniformly, different parts of the spectrum scale very differently. Thus

1. At $\omega = \frac{2^{p+1}}{T_{n+1}}$, the actual new spectral components characterizing the $n+1$ -st level of period doubling,

$$\hat{x}_{(n+1)}(2^{p+1}) \approx \frac{1}{2\alpha} \left(\frac{1}{\alpha} - (-1)^p i \right) \hat{x}_{(n)} \left(\frac{2^{p+1}}{T_{n+1}} \right)$$

or

$$\left| \hat{x}_{(n+1)}(2^{p+1}) \right| \approx \frac{1}{2} \sqrt{\frac{1}{\alpha^4} + \frac{1}{\alpha^2}} \left| \hat{x}_{(n)} \left(\frac{2^{p+1}}{T_{n+1}} \right) \right| \quad (35)$$

That is, the new spectral components are obtained by scaling the previous interpolation at these new positions by a factor of

$$\frac{1}{2} \sqrt{\frac{1}{\alpha^4} + \frac{1}{\alpha^2}} \approx \frac{1}{4.6}$$

or, dropped logarithmically by

$$10 \log_{10} 4.6 \approx 6.6$$

(The approximation (35) is the same as that for the scaling of successive RMS averages of spectral lines as obtained by Nauenberg [7].

2. At $\omega = \frac{2^{p+1}}{T_n}$, that is, at those frequencies of the n -th level, one finds the $n+1$ interpolation scaled by

$$\hat{x}_{n+1} \left(\frac{2^{p+1}}{T_n} \right) \approx \frac{1}{2\alpha} \left(\frac{1}{\alpha} + 1 \right) \hat{x}_{(n)}(2^{p+1}) \quad (36)$$

That is, in the next generation of doubling, the interpolation scales by the anomalously small amount of

$$\frac{1}{2\alpha} \left(\frac{1}{\alpha} + 1 \right) \approx \frac{1}{3.6}$$

or

$$10 \log_{10} 3.6 \approx 5.5 \text{ db}$$

at those frequencies that had just priorly come into existence.

3. At $\omega = \frac{2p}{T_n}$, one has

$$\hat{x}_{n+1} \left(\frac{2p}{T_n} \right) \approx \frac{1}{2\alpha} \left(\frac{1}{\alpha} - 1 \right) \hat{x}_n \left(\frac{2p}{T_n} \right) \quad (37)$$

where

$$\frac{1}{2\alpha} \left(\frac{1}{\alpha} - 1 \right) \approx - \frac{1}{8.3}$$

or

$$10 \log_{10} 8.3 \approx 9.2 \text{ db} .$$

Thus, in the second and all following generations after a spectral line has appeared, the successive interpolations drop anomalously quickly. (The approximation (37) is easily seen to be the same approximation of Grossman [8] for the "leading" edge of the spectrum).

4. In consequence of 2. and 3., if the n -th interpolation is raised by x for any $5.5 \text{ db} < x < 9.2 \text{ db}$, these spectra will all have regions of overlap. In particular, the original spectral prediction [4] of $x = 8.2 \text{ db}$ is included in this range. The present formula (34) is the full realization of the ideas of that previous paper.

5. Since different parts of the spectrum have different geometric scalings, a geometric mean of the n -th level spectral lines is a more significant average than the mean of the squares. Given (34), we can then compute the scaling of the geometric mean, or logarithmically, the average of log-amplitudes:

$$\hat{\chi}_n \equiv \frac{1}{2^{n-1}} \sum_{p=0}^{2^{n-1}-1} \ln |\hat{\chi}_{(n)}(2p+1)| \approx \int_0^1 dw \ln |\hat{\chi}_n(w)| .$$

Then,

$$\hat{\chi}_{n+1} - \hat{\chi}_n \approx \int_0^1 dw \ln \left| \frac{\hat{\chi}_{n+1}(w)}{\hat{\chi}_n(w)} \right| \approx -\ln(2\alpha) + \int_0^1 dw \ln \left| \frac{1}{\alpha} - e^{\pi i 2^n w} \right| .$$

However, the integral on the right identically vanishes for $|\alpha| > 1$. Thus,

$$\hat{\chi}_{n+1} - \hat{\chi}_n \approx -\ln(2\alpha) \quad (38)$$

or, the mean log amplitudes drop by

$$10 \log_{10}(2\alpha) \approx 10 \log_{10} 5.0 = 7.0 \text{ db} . \quad (39)$$

(This result is unchanged if the averaging is performed over all the n -th level spectral lines up to any multiple of the original fundamental, rather than just the subharmonic part.)

Equation (38) is a new result, and this approximate value has been numerically verified to be correct to the precision of (39).

At this point we want to extend these results to spectral properties at a fixed parameter value, rather than at successive period-doubling values. As previously demonstrated, as the parameter increases, the n -th level lines slightly increase until they saturate after several further period doublings. This increase is uniform for each level of introduced spectral lines from which it follows that properties 1.-5. are equally valid within a spectrum at fixed λ for all intermediate values of n (i.e., after several period doublings have occurred, but not for the lowest lying interpolations which have not yet saturated.) Taking the fundamental period to be 1, $w = 1$ is the original fundamental, and the n -th level of spectral lines are located at the frequencies

$$w = \frac{2p+1}{2^n} . \quad (40)$$

Now, all the spectral properties are consequences of (34) which itself is a consequence of (23). Working backwards, if we define, at a fixed parameter value λ_n , $d_{n,m}(t)$ as that part of $x_n(t)$ constructed purely from the m -th level spectral lines, then (23) must again hold for a new scaling function σ which has the same approximate value as (25). That is,

$$d_{n,m+1}(t) \sim \tilde{\sigma}(t/T_{m+1}) d_{n,m}(t) \quad (41)$$

where (41) is valid for all m such that $1 \ll m \ll n$ and $\tilde{\sigma}$ has the approximate value (25). Once $d_{n,m}(t)$ is specified, (41) is now a direct time-domain test of the theory. It is probably the best test to perform since just one time series is required, and there are no phasing problems. All that remains to be specified is $d_{n,m}(t)$:

$$d_{n,m}(t) = \frac{1}{2^{n-m+1}} \sum_{r=0}^{2^{n-m+1}-1} (-1)^r x_n(t+rT_{m-1}) \quad (42)$$

where

$$x_n(t) = \sum_{m=1}^n d_{n,m}(t) + \frac{1}{2^n} \sum_{r=0}^{2^n-1} x_n(t+rT_0) \quad (43)$$

so that $d_{n,m}$ is just that part of x_n which determines its m -th level spectrum.

For numerical experiments (41) is very readily verified; the theory leading to (41) will be published elsewhere. (41) can be iterated so that (41) and (43) approximately determine $x_n(t)$ from that part of it with period $2T_0$. This is a general structure arising from an underlying Cantor set, so that this type of data processing might be more generally applicable to dynamical systems with a strange attractor present. The analogue to (34) is

$$\hat{x}_{n,m+1}(\omega) \simeq \frac{1}{2\alpha} \left(\frac{1}{\alpha} - e^{\pi i \omega T_m} \right) \hat{x}_{n,m}(\omega) \quad (44)$$

so that the log-amplitude spectrum is built out of sums of determined periodic functions of successively doubled periods.

REFERENCES

- [1] D. Ruelle, *Comm. Math. Phys.* 55 (1977).
- [2] M. Misiurewicz, *Studia Math.* 67 (1980).
- [3] M. J. Feigenbaum, *J. Stat. Phys.* 21, 669 (1979).
- [4] M. J. Feigenbaum, *Phys. Lett.* 74A, 375 (1979), *Comm. Math. Phys.* 77, 65 (1980).
- [5] B. A. Huberman, J. Rudnick, *Phys. Rev. Lett.* 45, 154 (1980).
- [6] This formula first appeared as the scaling for the broadband part of a noisy spectrum in A. Wolf, J. Swift. University of Texas at Austin preprint "Universal Power Spectra for the Reverse Bifurcation Sequence" (1981).
- [7] M. Nauenberg, J. Rudnick, "Universality and the power spectrum at the onset of chaos." Preprint UCSC (1981).
- [8] S. Grossman, S. Thomae. University of Marburg, preprint (1981).
- [9] P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems.* (Birkhaeuser, Boston 1980).

TRICRITICAL POINTS AND BIFURCATIONS IN A QUARTIC MAP

Shau-Jin Chang, Michael Wortis, Jon A. Wright

Department of Physics
University of Illinois at Urbana-Champaign
1110 West Green Street
Urbana, IL 61801

We have studied the 1-dimensional iteration map associated with the even quartic polynomial $x_{n+1} = 1 + a x_n^2 + b x_n^4$. This map allows a smooth transition from a single hump to a double hump. Bifurcations and higher-order transitions occur as we vary the parameters a, b . In addition to the usual universal bifurcation behavior discovered by Feigenbaum, we find a new universality class of bifurcations which is associated with a tricritical point. Tricritical points serve as natural boundaries to Feigenbaum critical lines. For the quartic map, the tricritical points which are the end-points of the original Feigenbaum line are $(a, b) = (0, -1.59490)$ and $(-2.81403, 1.40701)$. Associated with each tricritical point, there are two unstable directions as well as two independent exponents. The exponents are $\delta_T^{(1)} = 7.2851$ and $\delta_T^{(2)} = 2.8571$. At the tricritical point, we can introduce a universal function $f_T^*(x)$ which obeys

$$\alpha_T f_T^*(f_T^*(x/\alpha_T)) = f_T^*(x)$$

with the scale factor $\alpha_T = -1.69031$. The quartic map has a special duality transformation $(a, b) \rightarrow (a', b')$, such that the two mappings are intrinsically related. The tricritical points which are dual to the above pair of tricritical points are located at $(-3.18980, 2.54371)$ and at $(.95561, -1.14981)$ and are joined by a line which is the dual of the original Feigenbaum line. There are an infinite number of different tricritical points which form at least a Cantor set.

I. INTRODUCTION

Recently, Feigenbaum has shown that the bifurcation sequence in a single hump iterative map $x_{n+1} = f(x_n)$ with a quadratic peak obeys a universal behavior.^{1,2} One can understand this behavior from a renormalization-group point of view. At the limiting point of a bifurcation sequence, the 2^N iteration of the map with appropriate scaling approaches a universal function $f_T^*(x)$. In the neighborhood of this universal function, it appears that there is only one relevant direction, along which the eigenvalue $\delta (= 4.6692)$ is larger than one. We have recently studied a more general map described by a quartic polynomial $x_{n+1} = f(x_n) \equiv 1 + a x_n^2 + b x_n^4$. An important property of the quartic map is that $f(x)$ can describe both a single-hump and a double-hump map. By changing the parameters a, b continuously, one can induce a smooth transition from a one-hump map to a two-hump map and vice versa. It is easy to see that a double-hump map admits many different kinds of stable cycles which do not exist in the iterations of a single-hump map. An important concept in a double-hump map is that of doubly-stable cycles. These are cycles whose members pass through both the central peak and a side peak. In the vicinity of a doubly-stable cycle in the a, b parameter space, both peaks control the

stability of the same cycle. Away from doubly-stable cycles, however, the peaks control different cycles which are unrelated in x -space. Thus, the regions controlled by these different peaks become dynamically coupled at a doubly-stable cycle. This kind of dynamical coupling is a new phenomenon which cannot occur in a single-hump map. Its presence modifies the bifurcation processes. The limiting point determined by a bifurcation sequence of 2^N doubly-stable cycles is associated with a fixed point in function space with two relevant directions in analogy with Feigenbaum's discussion of the quadratic case. This limiting point also serves as the end-point of a critical line which describes the usual Feigenbaum behavior. This type of fixed point is known as a tricritical point in phase-transition language.³ The tricritical point discovered here has two universal exponents $\delta_T^{(1)} = 7.2851$, $\delta_T^{(2)} = 2.8571$. At the fixed point, the map is self-similar near the peaks after 2^N iterations. Just as in the Feigenbaum's case, it is possible to introduce a universal function $f_T^*(x)$ at a tricritical point which obeys the same functional equation

$$\alpha_T f_T^*(f_T^*(x/\alpha_T)) = f_T^*(x)$$

but with a different scale factor $\alpha_T = 1.69031$. Our $f_T^*(x)$ can be expressed as a power series in x^4 . We have discovered an identity, $\delta_T^{(2)} = \alpha_T^2$, relating one of the exponents to the scale factor. This identity is a member of a whole class of identities which exist at the fixed point of an iterative map.

II. QUARTIC MAPS

(A) Single vs multiple-hump maps

We study the one-dimensional iterative maps generated by an even quartic polynomial

$$x_{n+1} = f(x_n) \equiv 1 + a x_n^2 + b x_n^4. \quad (2.1)$$

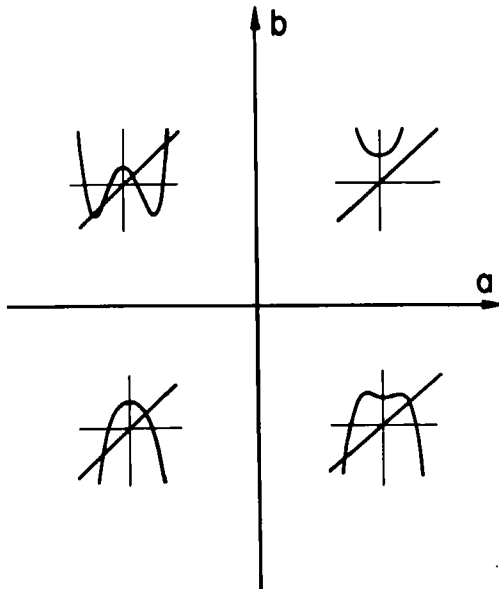


Figure 1

Iterative maps generated by the even quartic polynomial (2.1). The map in each quadrant corresponds to a $f(x)$ with (a, b) in that quadrant.

Depending on the values of a and b , $f(x)$ can have either a single hump or a double hump. See Fig. 1. For region defined by $a > 0$, $b > 0$, $f(x)$ can at most give rise to a stable fixed point. For $a < 0$, $b < 0$, the function $f(x)$ has a single quadratic hump, and Feigenbaum universality holds. In one parameter single-peaked maps, the limit point of infinitely bifurcated 2^N cycles is a single point. In two parameters single-peaked maps that point becomes a line that we refer to as the Feigenbaum line. For the regions $a < 0$, $b > 0$ and $a > 0$, $b < 0$, $f(x)$ has either two peaks and one valley or vice versa. In the following, we shall refer to the peak or valley at $x = 0$ as the central peak, and the peaks or valleys at $x \neq 0$ as the side peaks. The existence of both central and side peaks implies that $f(x)$ may develop independent iteration regions as indicated in regions I and II in Fig. 2.

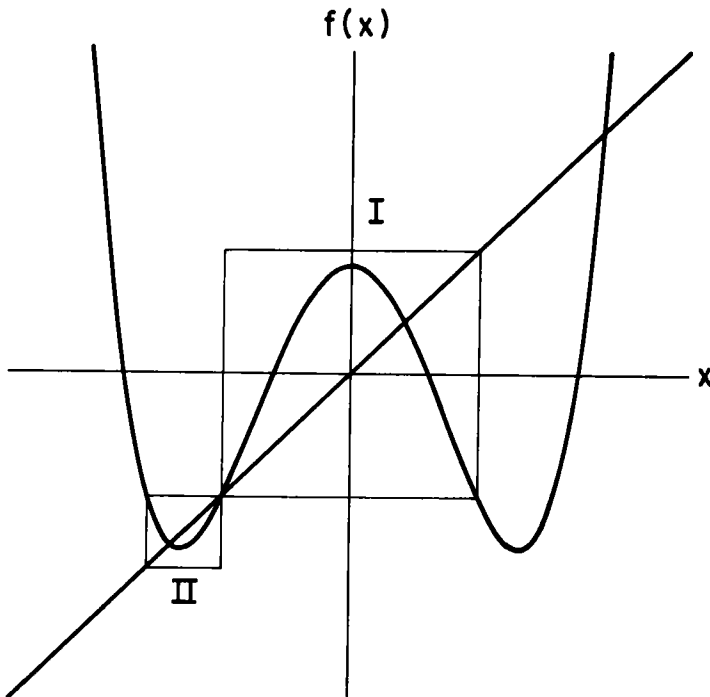


Figure 2

A map for $a < 0$, $b > 0$. Note that regions I and II are independent iteration regions.

In region I, $f(x)$ describes a single-hump map with a quadratic peak at $x = 0$. We shall refer to region I as the region controlled by the central peak. In region II, $f(x)$ describes an inverted single-hump map with the peak at $x = \sqrt{-2a/b}$. We shall refer to II as the region controlled by the side peak. Since the two side peaks have equal height, they map into the same x ($=1 - a^2/4b$). Hence, they cannot both be members of an N -cycle. Thus, we have at most one most stable region controlled by the side peaks. The existence of two independent regions may also appear in the case $a > 0$, $b < 0$.

(B) Duality transformation

For the study of cycles and their sequences, we only need to consider the region controlled by the central peak. The cycles associated with the side

peaks can be obtained from those with the central peak by a duality transformation described below.

Consider the quartic map (2.1). We can rewrite it as

$$\begin{aligned} x_{n+1} &= 1 + a x_n^2 + b x_n^4, \quad b \neq 0 \\ &= 1 - \frac{a^2}{4b} + b \left(\frac{a}{2b} + x_n^2 \right)^2. \end{aligned} \quad (2.2)$$

Thus, we can factor the quartic map as the product of two quadratic maps:

$$\xi_n = \frac{a}{2b} + x_n^2, \quad (2.3)$$

$$x_{n+1} = 1 - \frac{a^2}{4b} + b \xi_n^2. \quad (2.4)$$

We can now visualize the original mapping as

$$x_1 \rightarrow \xi_1 \rightarrow x_2 \rightarrow \xi_2 \rightarrow x_3 \rightarrow \dots. \quad (2.5)$$

It is easy to see that if the x -mapping has a stable N -cycle, so does the ξ -mapping and vice versa. The mapping from ξ_n to ξ_{n+1} is simply

$$\xi_{n+1} = \frac{a}{2b} + \left(1 - \frac{a^2}{4b} + b \xi_n^2 \right)^2. \quad (2.6)$$

After a rescaling, we can put the ξ -mapping (2.6) into the standard form

$$\xi_{n+1} = 1 + a' \xi_n^2 + b' \xi_n^4 \quad (2.7)$$

with

$$a' = \frac{(4b-a^2)[8ab + (4b-a^2)^2]}{32b^2} \quad (2.8)$$

and

$$b' = \frac{[8ab + (4b-a^2)^2]^3}{(8b)^4}. \quad (2.9)$$

It is easy to verify that the side peaks of x are mapped into the central peak of ξ , and the side peaks of ξ are mapped into the central peak of x . The duality relation is reciprocal: The duality transformation of (a', b') is the original (a, b) .

(C) Most-stable cycles and doubly-stable cycles

For a quartic map, the most-stable N -cycles (x_1, x_2, \dots, x_N) are described by

$$f(x_i) = x_{i+1}, \quad x_{N+1} = x_1, \quad i = 1, 2, \dots, N \quad (2.10)$$

and

$$\frac{d}{dx} [f^N(x)]_{x_i} = 0. \quad (2.11)$$

An N -cycle is most stable, if either the central peak ($x = 0$) or one of the side peaks ($x = \pm\sqrt{a/2b}$) is a member of the cycle x_i . The polynomials associated with a stable N -cycle give rise to a line in the a - b plane. In Fig. 3, the solid lines represent the loci of most-stable 2-cycles. Functions whose parameters satisfy $a + b + 1 = 0$ have most-stable 2-cycles associated with the central peak. The curved lines are the loci of points of most-stable 2-cycle associated with the side peaks. It is easy to see that, with the exception of the origin, the most stable N -cycle trajectories associated with the central peak never intersect among themselves. Neither do the most stable trajectories associated with the side peaks. However, the trajectories associated with the central peak do intersect those associated with the side peaks. Actually, they intersect in two completely different ways. At inter-

sections A and A' in Fig. 3, the most stable 2-cycles pass through both the central peak and one of the side peaks. We call this kind of cycle a doubly-stable cycle. Two trajectories which intersect at a doubly-stable cycle are "dynamically coupled" at the intersection. At intersections B and B', the 2-cycles are controlled by different peaks which are not related: some regions of x are attracted to one cycle and some to the other. We refer to this kind of intersection as "dynamically independent". Note that doubly stable points A and A' are dual to each other. The concept of dynamical coupling is invariant under duality transformation.

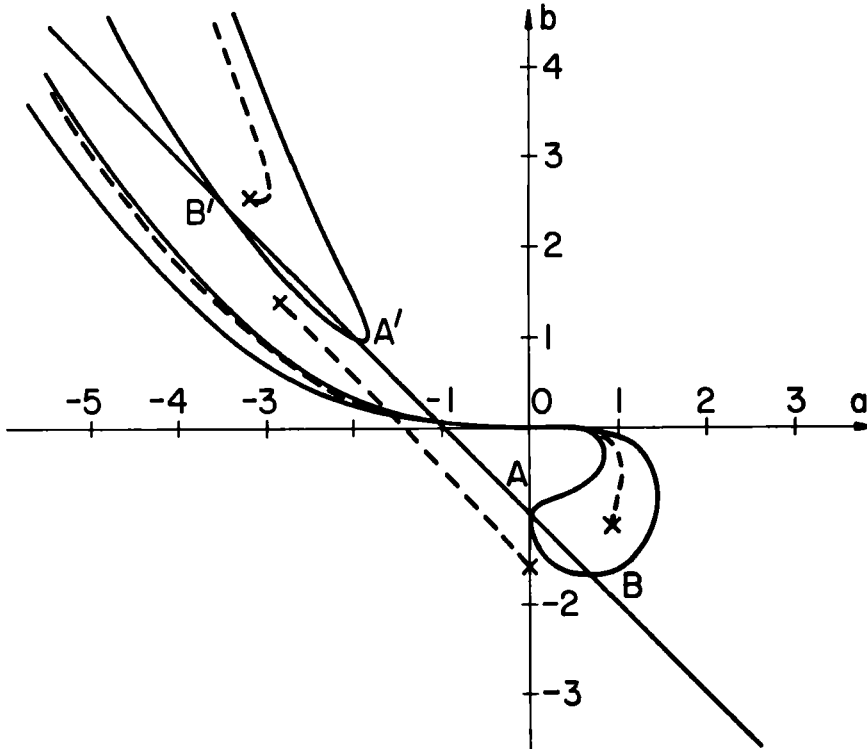


Figure 3

The solid lines represent the most-stable 2 cycles, the crosses denote tricritical points, and the dashed lines describe the Feigenbaum lines.

III. TRICRITICAL POINTS

(A) Bifurcation along b -axis

If we vary b (negative), keeping a small and negative, we encounter the usual bifurcation sequence, studied by Feigenbaum. If we vary b (negative), keeping a small and positive, we find that the map goes through three different most stable 2-cycles before bifurcating into a 4-cycle. These 2-cycles are controlled by the three different peaks as shown in Fig. 1. These phenomena persist no matter how small the value of a . The same phenomena repeat themselves at higher 2^N cycles. Thus, there are qualitative differences between bifurcations on the two different sides of the b -axis. When $a = 0$ and $b < 0$, the curve $f(x)$ is single-humped with a quartic maximum. We find a sequence of bifurcations which always occur at the doubly-stable 2^N cycles:

2^N	b_N
2	-1
4	-1.50393
8	-1.58225
16	-1.59316
32	-1.59466
...	...

The 2^N cycles have a limiting point at $b_\infty = -1.59490$, in close analogy to the quadratic map studied by Feigenbaum. However, because our map has a quartic (rather than quadratic) maximum, it is to be expected that there will be different critical exponents,⁴ as we now discuss. This limiting point has several interesting properties which we shall discuss below.

(B) Critical exponents and the universal function

There are two relevant eigen-directions at the critical point $(0, -1.59490)$. Along a relevant eigen-direction and for large N , we have

$$b_N = b_\infty + \frac{\text{const}}{\delta_T^N} \quad \text{with } |\delta_T| > 1 \quad (3.1)$$

where δ_T is the critical eigenvalue along this eigen-direction. One eigen-direction is along the b -axis where $\delta_T^{(1)} = 7.2851$. The other eigen-direction is $(1, -1.2347)$ where $\delta_T^{(2)} = 2.8571$. Note that the Feigenbaum fixed point has only one relevant direction with $\delta_F = 4.6692$. In phase-transition language, a critical point with two relevant directions which also serves as the end-point of a critical line is known as a "tricritical" point. As we shall see, our fixed point is indeed a tricritical point. See Discussion Section below.

At the critical point $(0, -1.59490)$, the function is self-similar near $x = 0$ to within a proper rescaling. The scale factor here is $\alpha_T = -1.6903$. Just as in the Feigenbaum case, the tricritical point has its own universal function $f_T^*(x)$ obeying the same functional equation

$$\alpha_T f_T^{*(2)}(x/\alpha_T) = f_T^*(x), \quad \alpha_T = -1.6903. \quad (3.2)$$

In the neighborhood of $x = 0$, $f_T^*(x)$ can be represented approximately by

$$f_T^*(x) = 1 - 1.83413 x^4 + 0.01301 x^8 + 0.31188 x^{12} - 0.062035 x^{16} + \dots \quad (3.3)$$

(C) An identity

The second exponent $\delta_T^{(2)}$ and the scale factor α_T are related by

$$\delta_T^{(2)} = \alpha_T^2. \quad (3.4)$$

This relation is exact,² as can be shown as follows: Let $h(x)$ be the eigenvector belonging to $\delta_T^{(2)}$. If

$$f(x) = f_T^*(x) + \epsilon h(x), \quad 0 < \epsilon \ll 1, \quad (3.5)$$

then iteration under

$$g(x) \equiv \frac{1}{f(1)} f^{(2)}(xf(1)) \quad (3.6)$$

leads by definition to

$$g(x) = f_T^*(x) + \delta_T^{(2)} \epsilon h(x) + O(\epsilon^2). \quad (3.7)$$

Since $h(0) = 0$ (because $f(0) = 1$), we may take

$$h(x) = x^2 + c_4 x^4 + c_6 x^6 + \dots \quad (3.8)$$

Direct iteration of (3.6) gives $(\alpha_T = 1/f_T^*(1))$

$$\begin{aligned} \alpha_T^{-1} \delta_T^{(2)} h(x) = & -h(1) f_T^*(x) + h\left(f_T^*\left(\frac{x}{\alpha_T}\right)\right) \\ & + \left[h(1) x f_T^{*'}\left(\frac{x}{\alpha_T}\right) + h\left(\frac{x}{\alpha_T}\right)\right] f_T^{*'}\left(f_T^*\left(\frac{x}{\alpha_T}\right)\right), \end{aligned} \quad (3.9)$$

which is a functional equation for $h(x)$. It is easy to verify that the right-hand side vanishes at $x = 0$. Only the last term contributes to $O(x^2)$ and by equating both sides of (3.9) one finds

$$\delta_T^{(2)} = f_T^{*'}(1)/\alpha_T. \quad (3.10)$$

One evaluates $f_T^{*'}(1)$ by differentiating (3.2) with respect to x and then setting $x = 0$. The result is

$$f_T^{*'}(1) = \alpha_T^3, \quad (3.11)$$

which gives (3.4). The other eigenvalues belonging to directions out of the x^4, x^8, x^{12}, \dots subspace are irrelevant. Eq. (3.4) is an example of a general class of identities which we discuss elsewhere.

(D) The tricritical point and doubly-stable cycles

Our present tricritical point $(0, -1.59490)$ is the bifurcation limit of a sequence of doubly-stable 2^N cycles. This turns out to be a general property of a bifurcation tricritical point. We have used this property to determine many other tricritical points which do not lie on the b -axis.

IV. DISCUSSION

In the $a < 0, b < 0$ region, we always have a single-hump function $f(x)$. If we increase the parameters $-a, -b$, we encounter an infinite sequence of bifurcations, as described by Feigenbaum. The limiting points of these infinite bifurcations now lie on a line in the a - b plane which we refer to as a Feigenbaum line. The tricritical point $(a, b) = (0, -1.59490)$ serves as a natural boundary to the original Feigenbaum line. When we extend the original Feigenbaum line in the other direction, we find that it terminates at another tricritical point, located at $(-2.81402, 1.40701)$. One can check easily that this tricritical point is also the limit of a sequence of doubly-stable 2^N cycles and that it has the same critical exponents $\delta_T^{(1)}$ and $\delta_T^{(2)}$.

Under the duality transformation, a tricritical point maps into a tricritical point and a Feigenbaum critical line maps into a Feigenbaum line. The dual images of the previous tricritical points are located at $(-3.18980, 2.54371)$ and at $(0.95561, -1.14981)$. We have plotted the dual transformed Feigenbaum line in Fig. 3.

Since bifurcation occurs after each stable cycle, and since there are an infinite number of different stable cycles, there are an infinite number of Feigenbaum lines in the a - b plane. Hence, there must also be an infinite number of tricritical points in the a - b plane. Indeed, we are able to show that between the critical points $(0, -1.59490)$ and $(.95561, -1.14981)$, there are an infinite number of tricritical points, forming a Cantor set. We shall discuss the geometrical meaning of the Cantor set elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant Nos. NSF PHY79-00272, NSF DMR78-21069, and by the Office of Naval Research under Contract No. N00014-80-C-0840.

REFERENCES

1. Feigenbaum, M. J., J. Stat. Phys. 19, 25 (1978); 21, 669 (1979), Phys. Lett. 74A, 375 (1979).
2. For reviews on iterative maps, see e.g. May, R. H., Nature 261, 459 (1976); Collet, P. and Eckmann, J. P., Iterated Maps on the Interval as Dynamical Systems (Birkhäuser Co., Boston 1980).
3. The term "tricritical point" was originally introduced by R. B. Griffiths: Griffiths, R. B., Phys. Rev. Lett. 24, 715 (1970). See also, Riedel, E. K., Phys. Rev. Lett. 28, 675 (1972); Griffiths, R. B., Phys. Rev. B7, 545 (1973).
4. Derrida, B., Gervois, A. and Pomeau, Y., J. Phys. A12, 269 (1979).

RECENT EXPERIMENTS ON CONVECTIVE TURBULENCE

Jerry P. Gollub

Physics Department, Haverford College, Haverford, PA 19041
and
Physics Department, University of Pennsylvania, Philadelphia, PA 19174

We have used laser Doppler methods to study the transition to turbulent convection in fluids of moderate Prandtl number. This work is briefly summarized, with emphasis on the major differences between large and small (relative to the depth) fluid layers. In small layers, we observe phenomena that are characteristic of nonlinear models with a few degrees of freedom, especially quasiperiodic oscillations, phase locking, and subharmonic generation. In large layers, the onset of chaotic motion is due to a structural instability of convection rolls, and low dimensional models are not likely to be useful.

INTRODUCTION

The problem of understanding the onset of turbulence in classical fluids is an old problem in nonlinear physics, but one that has attained new life due to recent advances in both mathematics and experiment. On the one hand, the hypothesis¹ that random behavior can arise from a strange attractor in a low dimensional dynamical system has held out a hope of new progress. On the other hand, recent experimental advances, such as laser Doppler velocimetry² and computer automation of experiments, have made a qualitative difference in the study of hydrodynamic instabilities and the transition to turbulence.

The Rayleigh-Benard instability has for many years played the role of the hydrogen atom of nonlinear hydrodynamics, because of its high degree of simplicity. Yet even a system as simple as a fluid layer with an imposed temperature gradient shows an amazing variety of interesting phenomena. During the past 5 years we have undertaken a detailed study of the processes leading to turbulent convection, with particular emphasis on evaluating the relevance of the strange attractor concept. Since this work has appeared (or will soon appear) elsewhere,²⁻⁸ we will summarize the main results and refer the reader to the literature for details. We do not attempt to review the experiments of other groups here.

Our experiments used water as the working fluid because its strongly temperature-dependent viscosity permitted the Prandtl number to be varied over a relatively wide range simply by varying the mean working temperature. (The Prandtl number is the ratio of the kinematic viscosity to the thermal diffusivity.) The rectangular fluid cells were designed with an emphasis on temperature stability and horizontal uniformity, because external perturbations can obscure the onset of intrinsic dynamical noise in the fluid, and nonuniformities can blur sharp transitions.

The fluid velocity at a point in the fluid was determined in real time from the Doppler shift of scattered laser light. Velocities as small as 10^{-4} cm/s could be determined with this technique. The fluid system was mounted on computer-

controlled stepping motor driven translation stages so that the velocity field could be mapped out in a horizontal plane, or alternately, studied as a function of time at a fixed location. A total of approximately 10^7 measurements were made in the course of this study.

The qualitative nature of the transition to turbulence is a strong function of the aspect ratio Γ , defined as the ratio of the largest horizontal dimension to the layer depth. The central result is that small aspect ratio fluid layers have many features in common with low dimensional nonlinear models, while large aspect ratio layers behave quite differently.

SMALL ASPECT RATIO

Small aspect ratio layers ($\Gamma \leq 5$) typically have a wide regime of time-independent convection when the critical Rayleigh number R_c (proportional to the temperature difference across the layer) is exceeded. Our Doppler mapping technique² facilitated the discovery that there are a number of stable convection patterns, depending on the past history or initial conditions of the system. For example, certain cells can contain either two or three convective rolls. Each pattern has a distinct and somewhat different sequence of instabilities leading to turbulence as R is varied, but each sequence is reproducible as long as the basic convection pattern remains unchanged. By varying the aspect ratio, Prandtl number, and convection pattern, we found several qualitatively distinct routes to turbulence in small ratio convection:

(a) Quasiperiodicity and phase-locking:^{2,4} The flow first becomes time-dependent by entering a strictly periodic regime at R_1 . At a higher Rayleigh number R_2 , a second frequency begins to grow smoothly as R is increased. These two oscillations interact with each other, causing the spectrum to consist of sharp peaks at the fundamental frequencies f_1 and f_2 , and all of their low order sums and differences of the form $f = m_1 f_1 + m_2 f_2$. The presence of these mixing components in the spectrum indicates that the time-dependent processes are strongly nonlinear. The ratio f_2/f_1 decreases smoothly with increasing R , indicating that the two frequencies are generally incommensurate, and hence that the spectrum corresponds to a quasiperiodic motion, which can be thought of as a torus in phase space. Usually, the two incommensurate oscillations phase-lock with each other, so that the ratio f_2/f_1 is the ratio of two small integers (typically $7/3$ or $9/4$). The spectrum then consists of an array of spikes at all multiples of the lowest commensurate frequency of f_1 and f_2 . The phase locking persists only over a relatively narrow range in R . When the ratio f_2/f_1 began to change again, the spectral lines became broadband, indicating the onset of nonperiodicity at R_t . At high R the line structure of the spectrum is no longer present, and its shape is similar to that found by Ahlers and Behringer⁹ at larger aspect ratio, with a power law falloff as $f^{-4.3 \pm 0.5}$. Nearer to the onset, the linewidth of the spectral peaks is a linear function of $(R - R_t)$. Phase locking has also been observed by Libchaber and Maurer.¹⁰

(b) Subharmonic (period doubling) bifurcations:^{2,7} Nonperiodicity can be obtained after several successive subharmonic bifurcations for certain combinations of aspect ratio and Prandtl number. This behavior is at least qualitatively similar to that shown by nonlinear mappings of the unit interval of the form $x_{n+1} = f(x_n)$. The initial time dependence is in the form of a periodic oscillation at frequency f . However, at a critical Rayleigh number R_1 peaks appear at $\frac{1}{2}f$ and odd harmonics, and at R_2 peaks at $\frac{1}{4}f$ and odd harmonics also appear in the spectrum. We have not resolved higher order subharmonics, but could have missed them by incrementing R by too large a step. The spectral peaks become broadened and a broad background begins to grow at a well defined value R_∞ . The strength of this background noise grows as $(R - R_\infty)^{1.43 \pm 0.2}$. This observation is consistent with a recent prediction by Huberman and Zisook¹¹ that the noise would grow as a power law with exponent $1.5247 \dots$. However, it is possible

that the agreement is fortuitous since we have not observed an extensive sequence of bifurcations. Subharmonics have been studied more extensively by Libchaber and Maurer¹⁰ and recently by Giglio, Musazzi, and Perini.¹²

(c) Three frequencies:² Quasiperiodic states characterized by three distinct frequencies were found in several cases. It is non-trivial to actually demonstrate that this is the case. Our spectra typically had over 20 statistically significant peaks, but they could all be fitted to linear combinations of three basic frequencies to within about one part in 10^4 . Comparable fits could not be obtained using two independent frequencies. These facts, combined with the observation that the ratios of the three frequencies vary smoothly with R , provide strong evidence that they are incommensurate with each other. Thus, this state can be represented by a 3-torus in phase space. Nonperiodicity evolves from this state at higher R .

Summarizing the behavior observed at small aspect ratio, we find a great variety of phenomena depending on the parameters of the system and the initial conditions. However, the prevalence of relatively simple phenomena (period doubling bifurcations, quasiperiodic motion, and phase-locking) is reminiscent of the behavior shown by simple nonlinear models. For example, systems of coupled electronic oscillators¹³ are known to exhibit phase locking. The following picture may assist the reader in organizing these various observations. Imagine a phase space that is peppered with limit cycles, tori, and strange attractors, each object surrounded by its own basin of attraction. As the Rayleigh number is varied, some of these objects shrink, while others grow. A comprehensive quantitative description of the entire phase space will probably be impossible to attain. On the other hand, the individual phenomena observed in each basin of attraction are reasonably well understood in principle.

LARGE ASPECT RATIO

When the largest horizontal dimension is much greater than the fluid depth, the behavior of the system is qualitatively different. Experiments by Ahlers and Behringer⁹ on cylindrical cells of convecting liquid helium showed that the strictly periodic and quasiperiodic regimes are absent and that the motion is chaotic whenever the velocity field is time-dependent. In fact, the existence of any regime of strictly time-independent convection was in doubt in these experiments. It is not possible to determine the cause of the early time dependence, because only the total heat flux through the system could be measured. We have recently extended our laser Doppler measurements to a rectangular cell of aspect ratio 30. Our scanning technique permits the structure of the velocity field to be mapped in the horizontal plane. This mapping could be done in a time short compared to the characteristic times of velocity fluctuations near their onset, so that the space and time structure of one component of the velocity field $\vec{v}(\vec{r}, t)$ could be recorded in digital form.

We found that the system is characterized by very long transients, comparable to the horizontal thermal diffusion time from one edge of the fluid layer to the other. This time is of the order of 10^5 s. In the interval $R_c < R < 4R_c$, the fluid attains a time-independent steady state after a day or two. However, this steady state is not the uniform one predicted by stability theory. Rather, the rolls preferentially align with their axes perpendicular to the walls of the cell, as shown in Fig. 1. The resulting pattern is splayed and contains defects, where a roll of positive (or negative) vorticity ends.⁸

Beyond about $4R_c$ (at Prandtl number 2.5), the local velocity begins to fluctuate slowly but irregularly. The corresponding power spectra are broadband, with a linewidth (square root of the second moment) that is approximately linear in $(R - R_t)$ above $R_t = 4R_c$. The linewidth is approximately the inverse of the

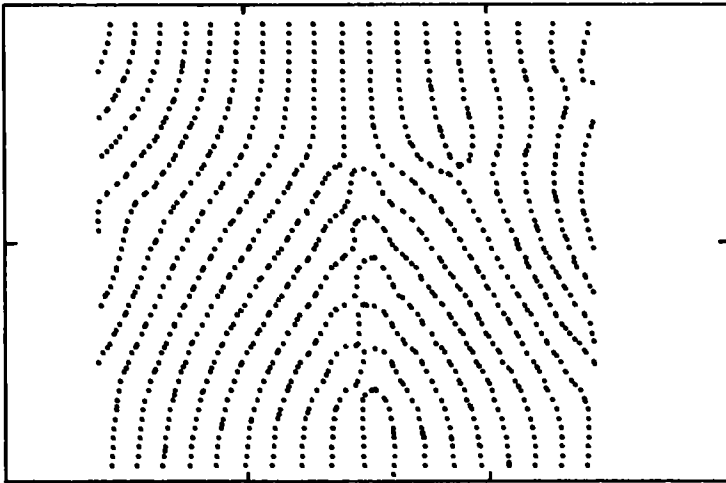


Fig. 1. Doppler map showing an example of the structure of the velocity field above the midplane of the $10 \times 15 \times 0.5$ cm cell in the time-independent steady state. The dots mark the roll boundaries, where the flow is entirely vertical (out of or into the page). The pattern contains defects and the rolls are perpendicular to the lateral boundaries.

characteristic time of velocity fluctuations. This time becomes quite long, about 7000 s at $5R_C$, thus permitting us to follow the space and time structure of the fluctuations in detail by repetitive laser Doppler imaging. We scanned a portion of the center of the cell every 10 minutes for runs lasting about 10 hours, obtaining 60 images per run in digital form which could be later reconstructed. Several sequences of these images are shown in Fig. 2. Shaded and unshaded regions are rolls of opposite vorticity (direction of circulation). At the boundaries between these regions, the fluid moves vertically (out of or into the page). In (b) the intrusion of a defect p is visible. In (c) it merges with a roll q of similar vorticity. This process forms a defect of opposite polarity (shaded) which is then expelled from the field of view. A second example begins in (f) where a roll labelled r is pinched off to form a defect, merges with region s in (h), and is then separated once again in (i).

These events indicate that the convective rolls are unstable with respect to deformations, and we believe that this instability is responsible for the onset of turbulence. The observed onset of the instability is consistent with the predicted onset of the "skewed varicose" instability of Busse and Clever.¹⁴ While the linear stability theory does not predict any time dependence, it is possible that nonlinear effects would cause this instability to evolve into chaotic time dependence. Thus, nonlinear analysis will be crucial in explaining the onset of turbulent convection at large aspect ratio. However, the strange attractor concept is not likely to be useful in this case.

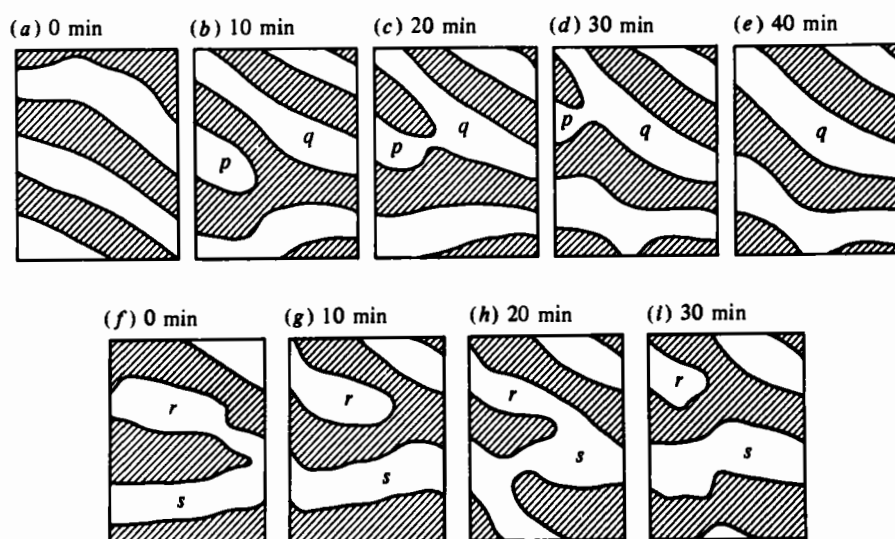


Fig. 2. Two sequences of Doppler images showing a portion of the velocity field at intervals of 10 minutes, at $R = 4.8R_C$. Shaded and unshaded regions have opposite circulation. The rolls are clearly unstable.

ACKNOWLEDGEMENT

This work was supported by a National Science Foundation Grant (CME 79-12150) to Haverford College. The experiments were performed in collaboration with J.F. Steinman and S.V. Benson.

REFERENCES

1. For a review, see Ruelle, D., *Strange Attractors*, La Recherche 108 (Fevrier 1980).
2. Gollub, J.P. and Benson, S.V., Many Routes to Turbulent Convection, *J. Fluid Mech.* 100 (1980) 449-470.
3. Gollub, J.P. and Benson, S.V., Chaotic Response to Periodic Perturbation of a Convecting Fluid, *Phys. Rev. Lett.* 41 (1978) 948-951.
4. Gollub, J.P. and Benson, S.V., Phase Locking in the Oscillations Leading to Turbulence, in: Haken, H. (ed.), *Pattern Formation by Dynamic Systems and Pattern Recognition* (Springer-Verlag, Berlin, 1979).
5. Gollub, J.P. and Steinman, J.F., External Noise and the Onset of Turbulent Convection, *Phys. Rev. Lett.* 45 (1980) 551-554.

6. Gollub, J.P., The Onset of Turbulence: Convection, Surface Waves, and Oscillators, in: Garrido, L. (ed.), *Systems Far from Equilibrium*, Lecture Notes in Physics Vol. 132 (Springer-Verlag, Berlin, 1980).
7. Gollub, J.P., Benson, S.V. and Steinman, J.F., A Subharmonic Route to Turbulent Convection, *Ann. New York Acad. Sci.* 357 (1980) 22-27.
8. Gollub, J.P. and Steinman, J.F., Doppler Imaging of the Onset of Turbulent Convection, to appear.
9. For example, see Ahlers, G. and Behringer, R.P., The Rayleigh Benard Instability and the Evolution of Turbulence, *Progr. Theor. Phys. Suppl.* No. 64 (1978) 186-201.
10. Libchaber, A. and Maurer, J., Une Experience de Rayleigh-Benard de Geometrie Reduite: Multiplication, Accrochage et Demultiplication de Frequences, *J. de Physique* 41 (1980) C3-51 to C3-53.
11. Huberman, B.A. and Zisook, A.B., Power Spectra of Strange Attractors, *Phys. Rev. Lett.* 46 (1981) 626-628.
12. Giglio, M., Musazzi, S. and Perini, U., Transition to Chaos via a Well Ordered Sequence of Period Doubling Bifurcations, preprint.
13. Gollub, J.P., Brunner, T.O. and Danly, B.G., Periodicity and Chaos in Coupled Nonlinear Oscillators, *Science* 200 (1978) 48-50.
14. Busse, F.H. and Clever, R.M., Instabilities of Convection Rolls in a Fluid of Moderate Prandtl Number, *J. Fluid Mech.* 91 (1979) 319-335.

EXPERIMENTAL OBSERVATIONS OF COMPLEX DYNAMICS IN A CHEMICAL REACTION

J.-C. Roux,* Jack S. Turner,[†] W. D. McCormick, and Harry L. Swinney

Department of Physics
The University of Texas at Austin
Austin, Texas 78712
U.S.A.

In experiments on a stirred flow chemical reactor we have observed a sequence of periodic and chaotic regimes that alternate as a function of the flow rate. The periodic regimes are characterized by power spectra consisting of a single fundamental frequency component and harmonics and by limit cycle attractors, while the chaotic regimes are characterized by broadband power spectra and strange attractors. One of the chaotic regimes has been studied in detail and has been found to correspond to a smooth one-dimensional map that has a single maximum and a positive Lyapunov exponent.

INTRODUCTION

Oscillating nonequilibrium chemical systems were observed unambiguously early in this century, but their existence remained largely unknown until recently, even among chemists. Now nonequilibrium chemical systems are being widely studied as examples of nonlinear systems which have interesting dynamics. Examples of nonequilibrium systems that often exhibit periodic or more complex oscillations include living systems and the chemical stirred flow reactors used in industry.

By far the most extensively investigated and best understood oscillating chemical system is the Belousov-Zhabotinskii (BZ) reaction, which was discovered [1] and studied [2] in the Soviet Union about 20 years ago. In this reaction an acidic bromate solution oxidizes an organic compound in the presence of a metal ion catalyst. The mechanism of the BZ reaction, which involves more than 20 reactions and as many chemical constituents, has been elucidated by Noyes and coworkers in a sequence of more than 30 papers [3-5]. On the experimental side, work by several groups, particularly Schmitz, Hudson, and coworkers [6-7] and the Bordeaux group [8-11], has revealed several types of dynamics in the BZ reaction, including multiple steady states, simple and complex oscillations, intermittency, and chaotic behavior.

We have conducted experiments on the BZ reaction in a stirred flow reactor and have observed a sequence of distinct dynamical regimes. The reaction was studied for different residence times τ (where τ = reactor volume/flow rate) with all other variables including the concentrations of input chemicals held fixed [12]. Note that the inverse residence time, τ^{-1} , like the Reynolds number in a hydrodynamic system, is a measure of the distance away from thermodynamic equilibrium for a chemical reaction maintained in a stationary regime.

Preliminary results of this work are presented here; detailed reports, including a numerical study of a model of the reaction, will be presented elsewhere [12-14].

EXPERIMENTAL SYSTEM

The chemicals were injected continuously with a multi-channel peristaltic pump into a vigorously stirred reactor that was immersed in a temperature-controlled bath. The experimental parameters were as follows: reactor volume, 31.0 cm³; bath temperature, 28.3°C; chemical concentrations in the mixed feed--0.25 M malonic acid, 0.14 M potassium bromate, 0.00083 M cerous sulfate octahydrate, and 0.2 M sulfuric acid; residence time, $0.5 < \tau < 4$ hr.

The concentration of one of the reaction products, bromide ion, was measured with a specific ion probe, digitized, and recorded as a function of time in a computer. The digital time series records $B(t_i)$ (where $t_i = i\Delta t$, $i = 1, \dots, 16384$, and $\Delta t = 0.88$ s) were Fourier transformed to obtain power spectra. In addition, the dynamical behavior of the system was characterized by phase space portraits constructed following a procedure suggested by Ruelle [15]: an n -dimensional (nD) phase portrait can be obtained by plotting successive points $[B(t_i+T_1), B(t_i+T_2), \dots, B(t_i+T_n)]$, where the T_j ($j=1, \dots, n$) are suitably chosen constants. (A related procedure, in which the nD portrait is obtained from a time series and its first $n-1$ derivatives, was proposed by Packard *et al.* [16] and tested in experiments on the BZ reaction by Roux *et al.* [17].) We report here some 2D portraits obtained with $T_1=0$ and $T_2=8.8$ s, and a Poincaré map obtained from a 3D portrait with $T_1=0$, $T_2=8.8$ s, and $T_3=17.6$ s.

EXPERIMENTAL RESULTS

For $\tau < 0.87$ hr and $\tau > 2.28$ hr the reaction exhibits simple periodic oscillations. When τ is increased within the range between these limits, the reaction passes through an alternating sequence of periodic and chaotic regimes; we call these regimes P_1 ($\tau < 0.87$ hr), C_1 , P_2 , C_2 , P_3 , C_3 , ..., P'_1 ($\tau > 2.28$ hr). The time dependence of the bromide ion potential (proportional to $-\log[Br^-]$) and the corresponding power spectra and 2D phase portraits are shown for the first six regimes in Figures 1-6. (The bromide potential is plotted in relative units with a zero offset; the potential oscillates in the range 155-200 mV.)

As Figures 1, 3, and 5 illustrate, each regime P_k is characterized by: (1) a periodic time series with each period containing one large amplitude relaxation oscillation (of the type in regime P_1) and $k-1$ small amplitude nearly sinusoidal oscillations (of the type in regime P'_1), (2) a power spectrum consisting of a single instrumentally sharp fundamental frequency component and its harmonics, and (3) a phase portrait that is a limit cycle.

As Figures 2, 4, and 6 illustrate, each regime C_k is characterized by: (1) a nonperiodic oscillating time series, (2) a broadband power spectrum, and (3) a phase portrait that is a strange attractor (as will be discussed in the following section). One way in which we have distinguished the periodic and chaotic regimes is by comparing zero-frequency intercepts of the background noise level in the power spectra for the different regimes. In the periodic regimes the broadband noise level, which is presumably due to instrumental noise (including fluctuations in flow rate, stirring rate, etc.), is typically two to three orders of magnitude lower than the broadband spectral intensity at low frequencies for the chaotic regimes; this is illustrated for P_2 and C_2 in Figure 7.

The sequence of dynamical regimes we have observed is summarized in Figure 8, which shows the broadband spectral noise level as a function of residence time. Each periodic and chaotic regime clearly exists over some range of τ , although the extent of the regimes, drawn as equal in Figure 8, has not been determined.

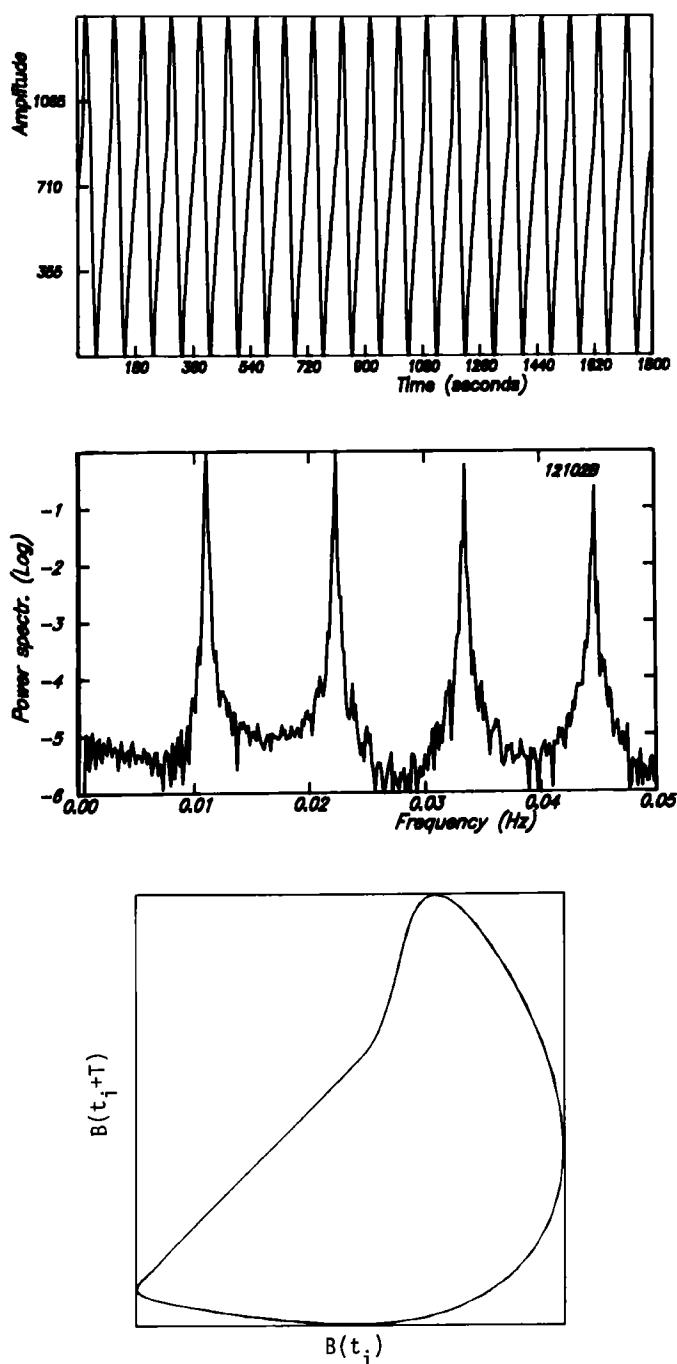


Figure 1. The time dependence of the bromide ion potential and the corresponding power spectrum and phase portrait (with $T = 8.8$ s) for data obtained in regime P_1 with $\tau = 0.49$ hr.

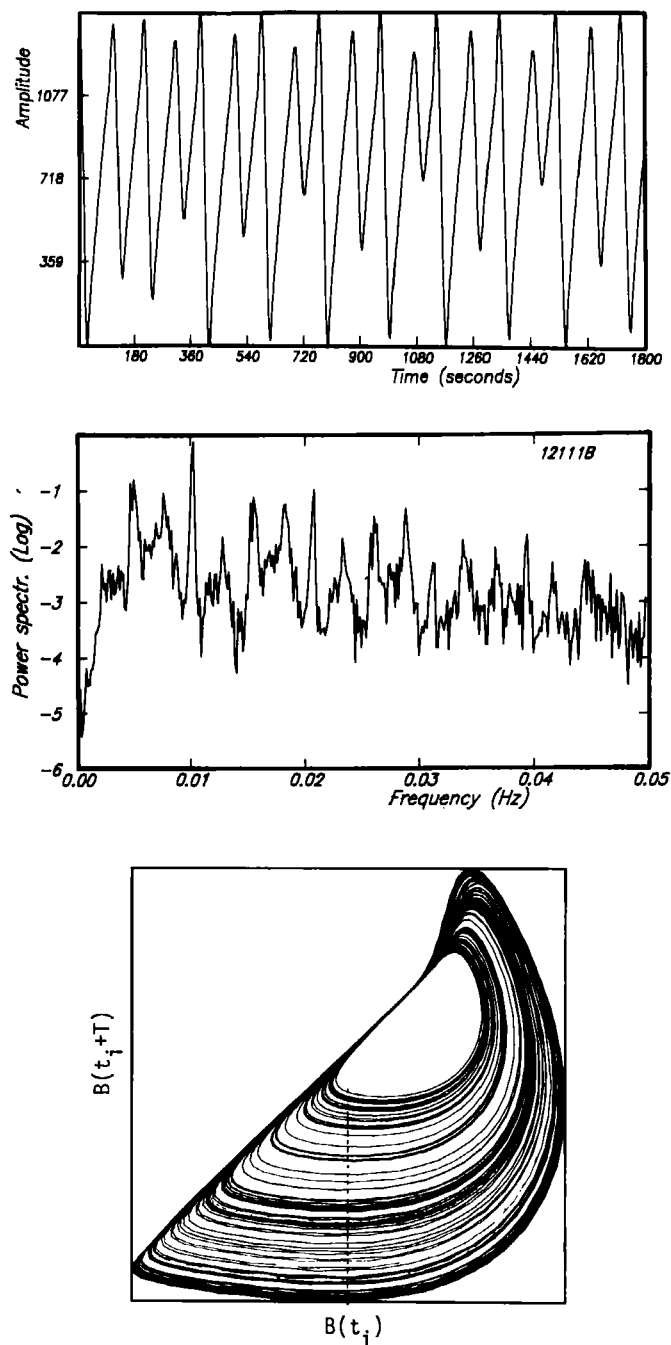


Figure 2. The time dependence of the bromide ion potential and the corresponding power spectrum and phase portrait (with $T = 8.8$ s) for data obtained in regime C_1 with $\tau = 0.90$ hr. The vertical dashed line shows the location of the Poincaré and return maps discussed in the section entitled Deterministic Chaos.

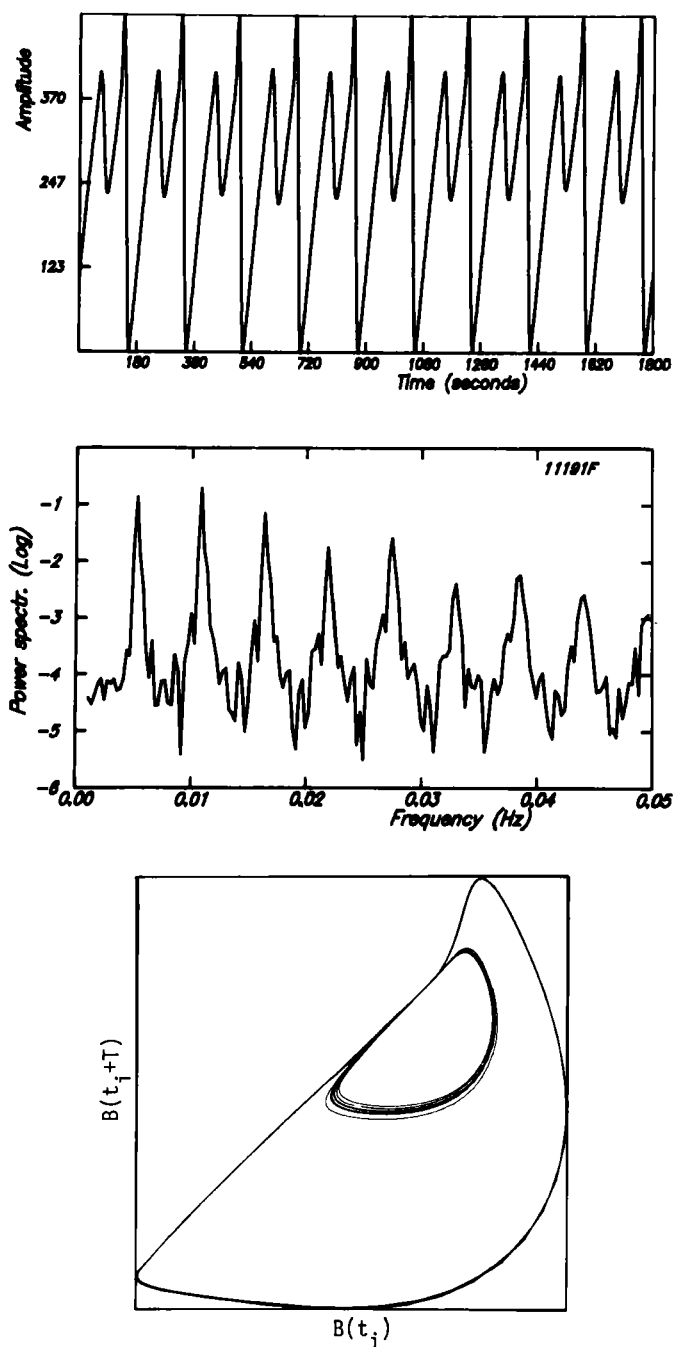


Figure 3. The time dependence of the bromide ion potential and the corresponding power spectrum and phase portrait (with $T = 8.8$ s) for data obtained in regime P_2 with $\tau = 1.03$ hr.

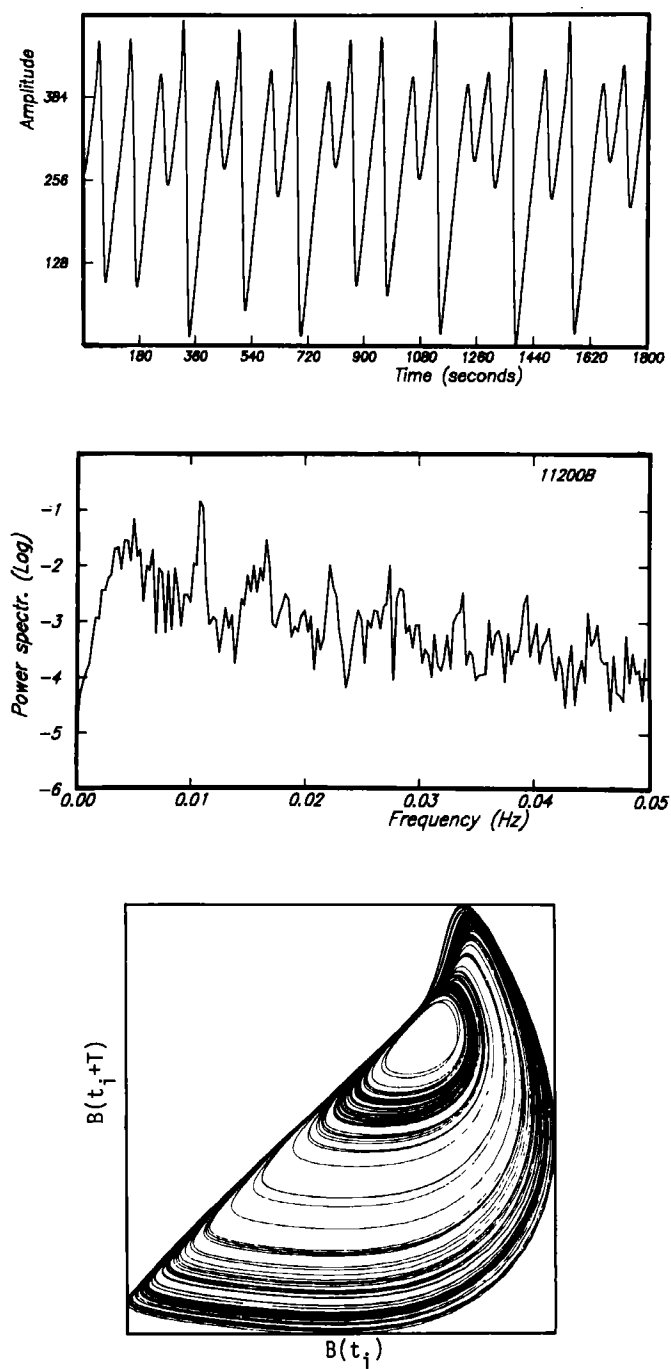


Figure 4. The time dependence of the bromide ion potential and the corresponding power spectrum and phase portrait (with $T = 8.8$ s) for data obtained in regime C_2 with $\tau = 1.11$ hr.

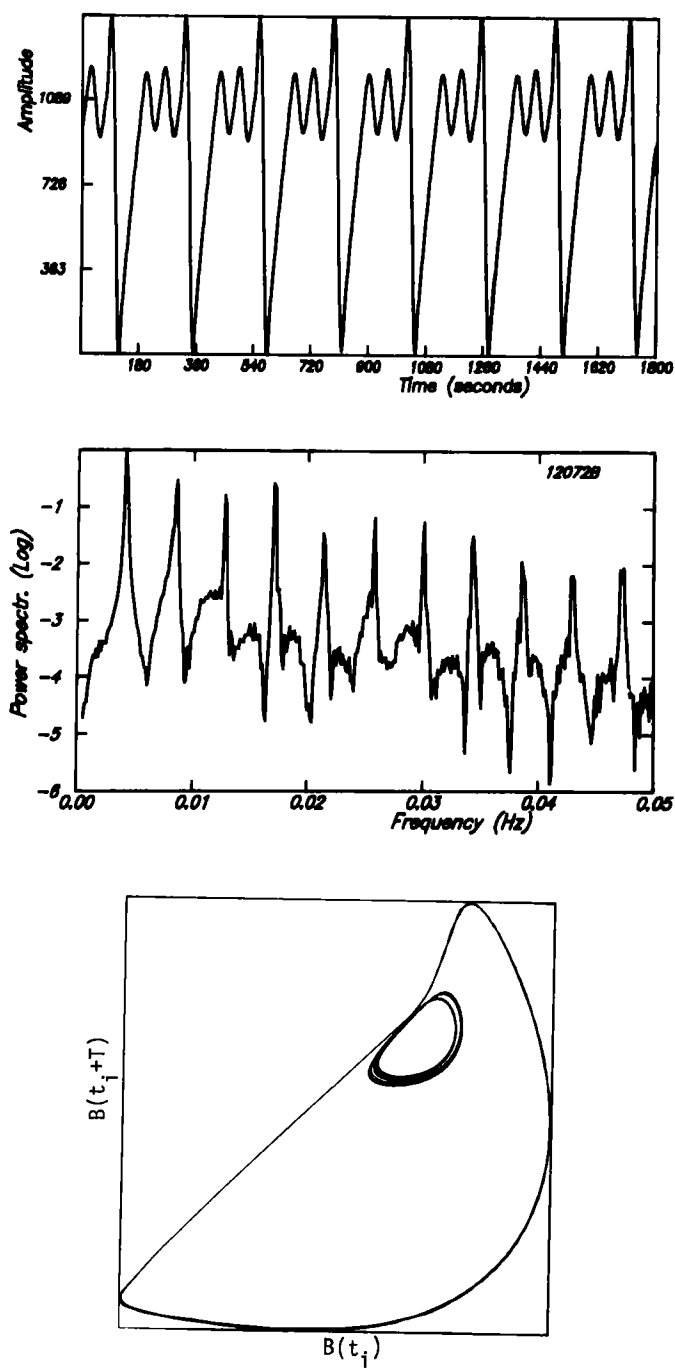


Figure 5. The time dependence of the bromide ion potential and the corresponding power spectrum and phase portrait (with $T = 8.8$ s) for data obtained in regime P_3 with $\tau = 1.25$ hr.

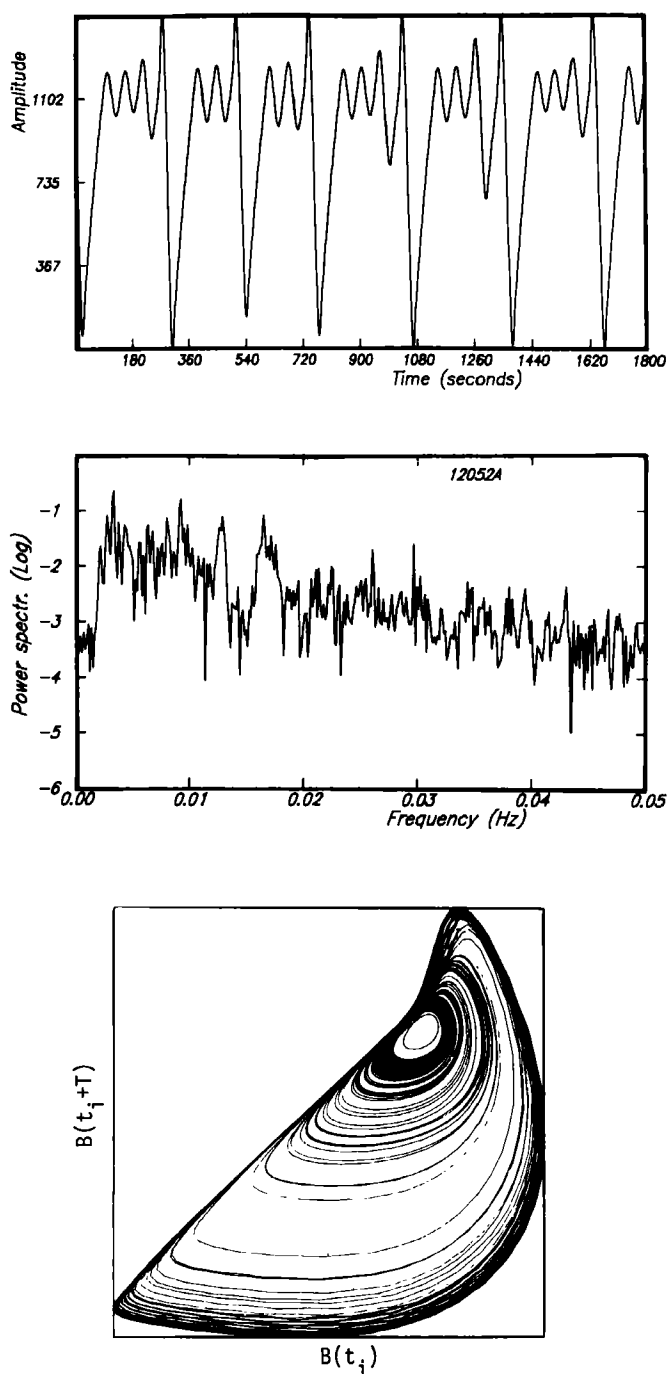


Figure 6. The time dependence of the bromide ion potential and the corresponding power spectrum and phase portrait (with $T = 8.8$ s) for data obtained in regime C_3 with $\tau = 1.38$ hr.

DETERMINISTIC CHAOS

We have called the attractors for the regimes C_k "strange attractors," which implies that the system is deterministic (see e.g., Lanford [18]). As Showalter et al. [19] rightly emphasize, all experimental systems are inevitably subject to external perturbations which result in nonperiodic behavior; "therefore, failure to observe exact repetition in a chemical stirred tank reactor cannot prove the system would be truly chaotic in the complete absence of perturbation". We concur completely. No experiment will ever be able to make an absolute distinction between stochastic and deterministic processes, and, in fact, such a distinction cannot even be made in principle: as Kac has pointed out, a stochastic process can always in principle be devised to describe any noisy data even when the data can be simply and elegantly described by a deterministic model [20].

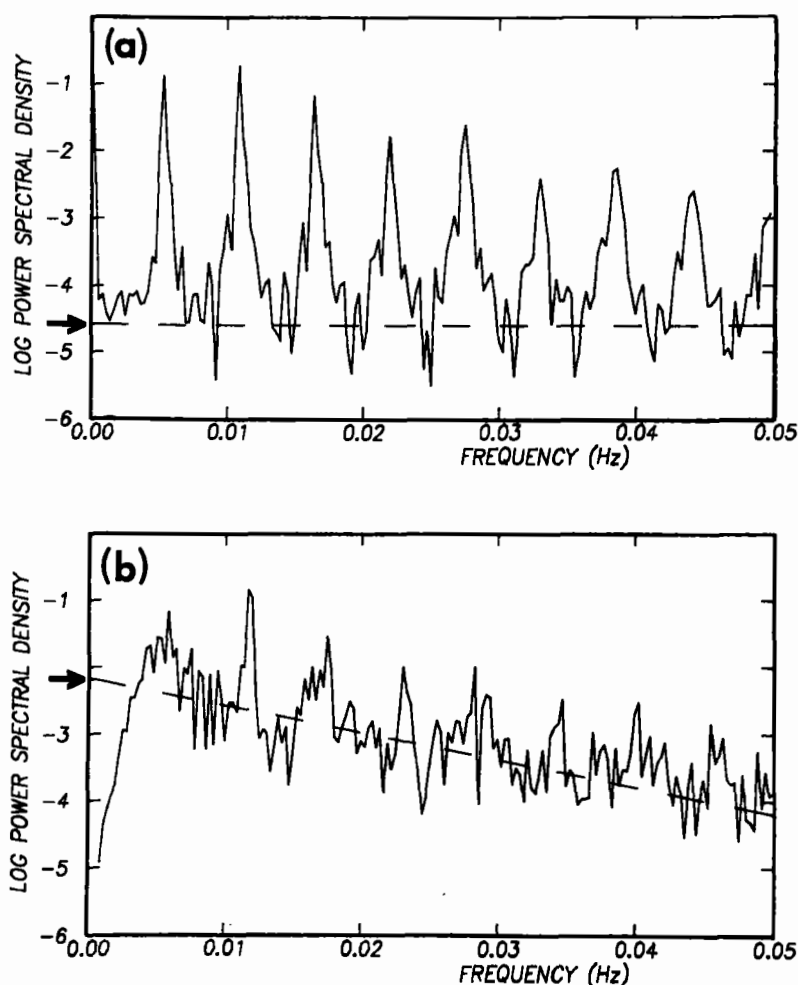


Figure 7. The intercepts at zero frequency of lines drawn through the broadband background noise are illustrated for (a) a periodic regime (P_2) and (b) a chaotic regime (C_2).

Recognizing these difficulties, we now present three data analyses that provide successively stronger support for the interpretation of regime C_1 as corresponding to deterministic chaos. First consider the 2D Poincaré map shown in Figure 9. This map was constructed from a 3D Ruelle-type portrait that has, in addition to the two axes shown in the phase portrait in Figure 2, a third axis (perpendicular to the first two axes) obtained with $T_3=17.6$ s. The Poincaré map in Figure 9 was formed by the intersection of the 3D phase space orbits with a plane that passes (normal to the paper) through the dashed line in Figure 2. The observation that the Poincaré map is a smooth curve (Figure 9) rather than a scatter of points suggests that the process is deterministic; however, a stochastic process could be devised to give the curve in Figure 9.

As a second test of the character of the dynamics, consider a return map constructed from the attractor in Figure 2. Let $X(n)$ be the value of the ordinate when an orbit in the 2D portrait crosses the dashed line in Figure 2, and let $X(n+1)$ be the ordinate the next time the orbit crosses the dashed line. A return map obtained by plotting the ordered pairs $[X(n), X(n+1)]$ for all n is shown in Figure 10. Again the data describe a smooth curve that suggests that the system is deterministic rather than stochastic: given any initial $X(n)$, all subsequent $X(n+k)$ are determined by the map.

The strongest evidence we have indicating that C_1 corresponds to deterministic chaos is given by the value of the Lyapunov exponent. Lyapunov exponents provide a quantitative measure of the degree of chaos or "sensitive dependence on initial conditions" for nonlinear systems. A positive exponent means that nearby orbits will exponentially separate; the system is then highly sensitive to the initial conditions. On the other hand, if the exponent is negative, nearby orbits will exponentially converge; the system is then insensitive to initial conditions.

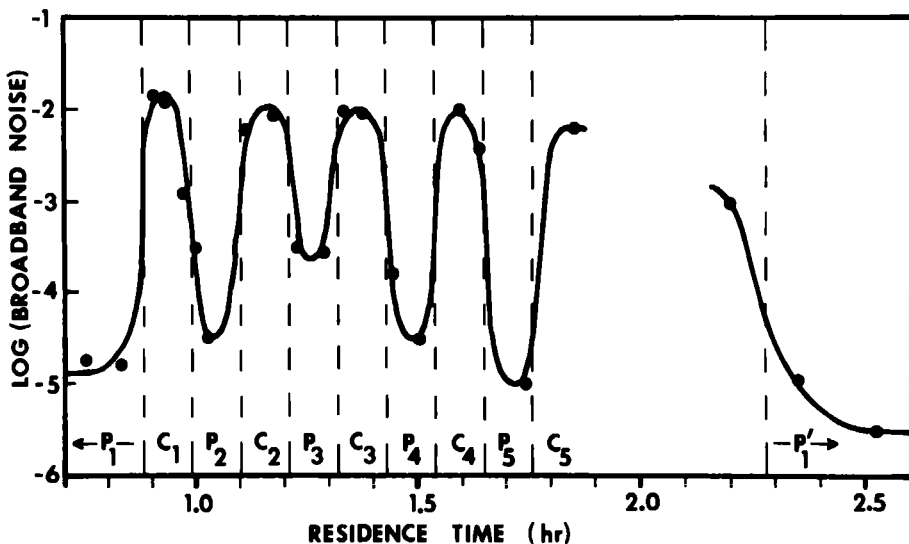


Figure 8. The alternating sequence of periodic and chaotic regimes is shown by this graph of the broadband spectral noise (zero-frequency intercepts) measured as a function of residence time. In the periodic regimes P_k the noise is instrumental (see the attractors in Figures 1, 3, and 5), while in the chaotic regimes C_k the noise is several orders of magnitude larger. The curve is drawn to guide the eye.

For a 1D map the Lyapunov exponent has a maximum value of $\ln 2 = 0.69$. Robert Shaw has calculated the Lyapunov exponent for the map shown in Figure 10 and has obtained 0.5 ± 0.1 . The positive Lyapunov exponent gives the rate of divergence of orbits in the attracting sheet shown in cross section in the Poincaré map (Figure 9). In contrast, orbits off of the sheet converge exponentially fast to the sheet (the Lyapunov exponent in directions normal to the sheet is negative). Therefore, the inevitable scatter in our experimental data due to fluctuations in the environment is much less noticeable in the Poincaré map (where the scatter is along the curve in Figure 9) than in the return map (Figure 10).

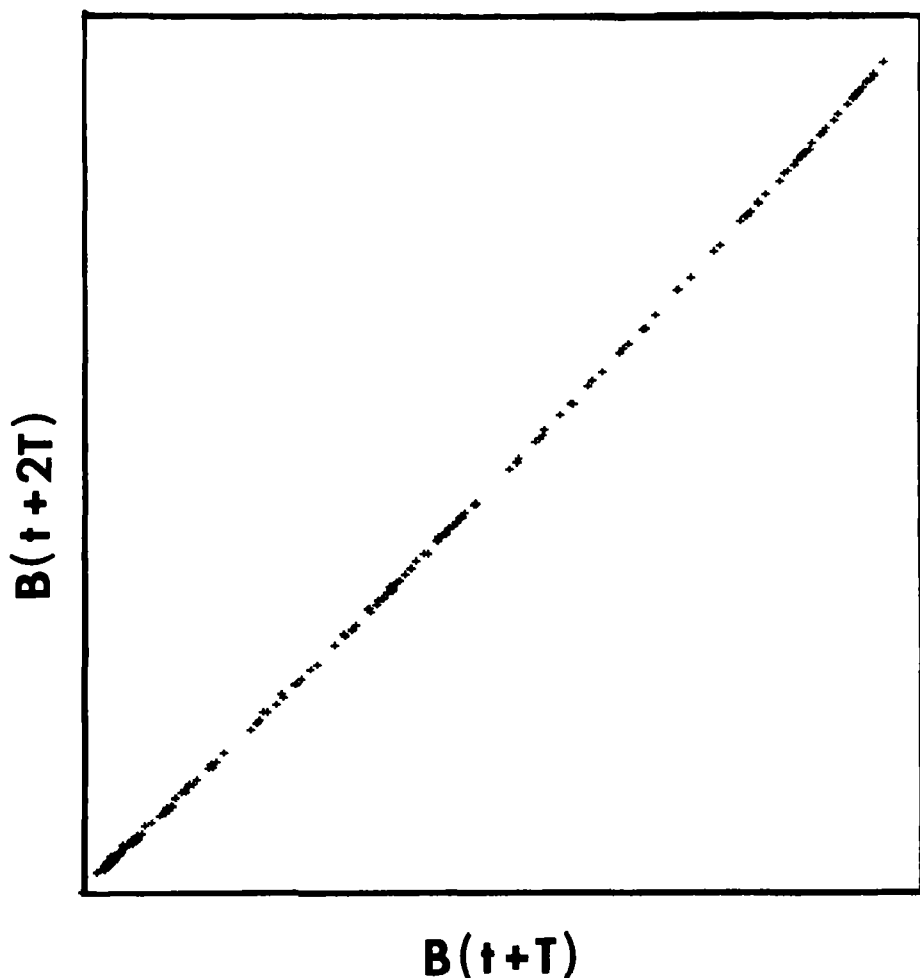


Figure 9. A Poincaré map (with $T = 8.8$ s) for the strange attractor in regime C_1 shown in Figure 2. The map was formed by the intersection of the 3D orbits with the plane (normal to the paper) passing through the dashed line in Figure 2.

MODEL STUDIES

Numerical studies of a mathematical model of the BZ reaction have played an important role in the design and guidance of our experimental program. In these studies a four-variable model based on the Oregonator of Field and Noyes [21,22] has been found to exhibit a variety of phenomena relevant to experimental studies in a stirred flow reactor. These include multiple stationary and oscillatory states, hysteresis, bursts of oscillation, complex periodic and nonperiodic states, and two distinct sequences of transitions involving mixed periodic and chaotic states. Details of the model and its predictions are presented elsewhere [12,14].

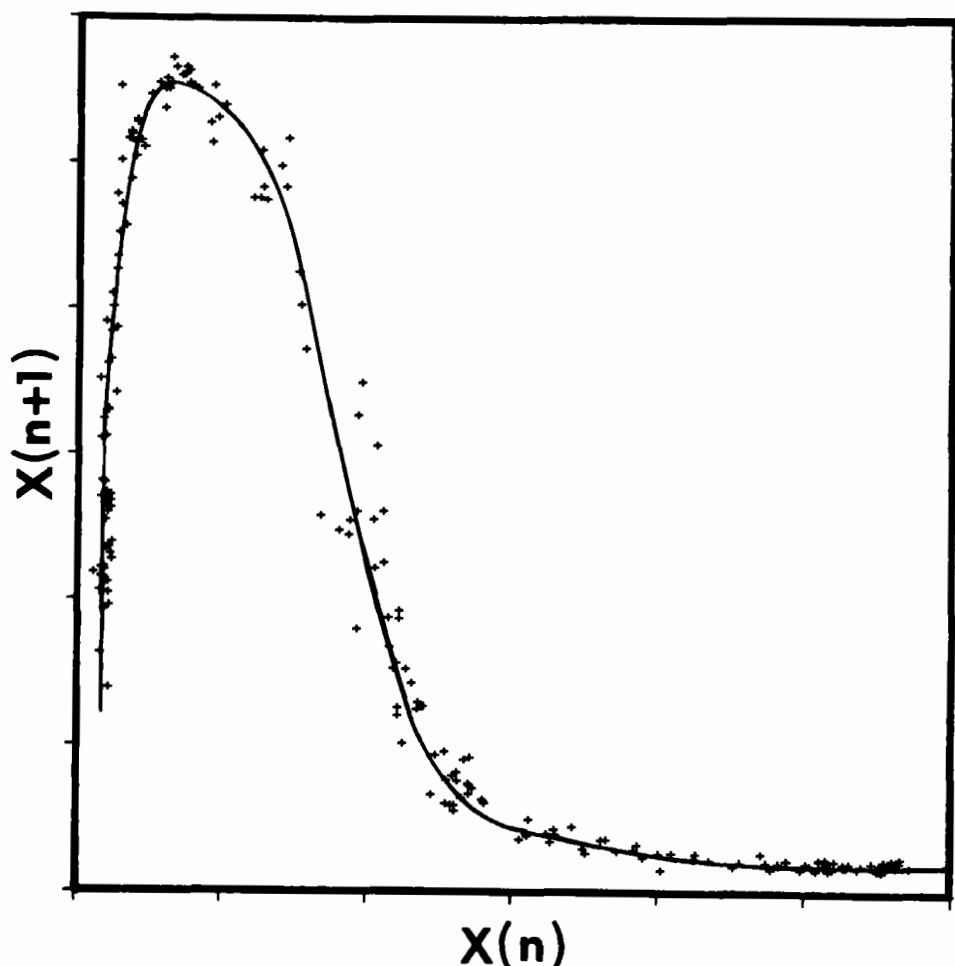


Figure 10. A return map for the strange attractor in regime C_1 shown in Figure 2. The map was formed by plotting as ordered pairs $[X(n), X(n+1)]$ the successive values $X(n)$ of the ordinate of phase space orbits when they cross the dashed line in Figure 2.

Of particular interest here are the sequences of transitions which occur as the inverse residence time is varied. The oscillations which appear first in the model as τ^{-1} increases are nearly sinusoidal. Further increase in τ^{-1} reveals a period-doubling sequence of transitions of the type discussed in this volume by Feigenbaum [see also Ref. 23]. For still larger τ^{-1} , a second type of oscillation appears in the model studies, having larger amplitude and lower frequency than those previously found. It is in this region of τ^{-1} that the periodic-chaotic sequence observed in our experiments is found in the model. The model predictions of the periodic-chaotic sequence, which were obtained prior to the conduct of the experiments, are in excellent qualitative agreement with the experiments [12]. We expect the quantitative agreement to improve with parameter variation and model refinement now in progress.

DISCUSSION

A similar sequence of alternating periodic and chaotic regimes has been observed in experiments by Hudson *et al.* [7] and by the Bordeaux group [10,24]. However, the sequences of regimes observed in our experiments [12] and in previous experiments [7,10,24] have the opposite dependence on residence time τ . The experimental conditions are similar except for the range in residence times studied: $0.08 < \tau < 0.2$ hr for previous experiments and $0.5 < \tau < 4$ hr for our experiments.

In addition to the numerical studies by Turner [12,14], the periodic-chaotic sequence has been found by Tomita and Tsuda [25] in a dynamical systems model (map) developed to describe the observations of Hudson. Thus the models and the experiments suggest that the periodic-chaotic sequence, like the period-doubling sequence, may occur widely in nonlinear systems.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with Robert Shaw and Jack Hudson. We thank Jack Hudson for a preprint [26], which we received after the present results were presented at the Los Alamos conference; the preprint describes an analysis of his experiments using phase portraits and 1D maps similar to those which we have presented here. This research is supported by National Science Foundation Grant CHE79-23627 and The Robert A. Welch Foundation Grant F-805.

*Permanent address: Centre de Recherche Paul Pascal, Université de Bordeaux-I, Domaine Universitaire, 33405 Talence Cedex, France.

*Also, Center for Studies in Statistical Mechanics, University of Texas, Austin, Texas 78712.

REFERENCES

- [1] Belousov, B. P., Ref. Radiats. Med. 1958 (1959) 145.
- [2] Zhabotinskii, A. M., Dokl. Akad. Nauk. SSSR 157 (1964) 392; Biofizika 9 (1964) 306.
- [3] Field, R. J., Körös, E., and Noyes, R. M., J. Am. Chem. Soc. 94 (1972) 8649.
- [4] Field, R. J. and Noyes, R. M., Accounts Chem. Res. 10 (1977) 214.
- [5] Noyes, R. M. and Field, R. J., Accounts Chem. Res. 10 (1977) 273.
- [6] Schmitz, R. A., Graziani, K. R., and Hudson, J. L., J. Chem. Phys. 67 (1977) 3040.
- [7] Hudson, J. L., Hart, M., and Marinko, D., J. Chem. Phys. 71 (1979) 1601.
- [8] De Kepper, P., Rossi, A., and Pacault, A., C. R. Acad. Sci. Paris 283C (1976) 371.
- [9] Vidal, C., Roux, J.-C., and Rossi, A., J. Am. Chem. Soc. 102 (1980) 1241.
- [10] Vidal, C., Roux, J.-C., Bachelart, S., and Rossi, A., N.Y. Acad. Sci. 357 (1980) 377.

- [11] Pomeau, Y., Roux, J.-C., Rossi, A., and Bachelart, S. and Vidal, C., submitted to *J. Phys. Lett.* (Paris).
- [12] Turner, J. S., Roux, J.-C., McCormick, W. D., and Swinney, H. L., submitted to *Phys. Lett.*
- [13] Swinney, H. L., Roux, J.-C., and King, G. P., in: Meijer, P. H. E., Mountain, R. D., and Soulen, R. J. (eds.), *Noise in Physical Systems* (Proceedings of the Sixth International Conference)(National Bureau of Standards Special Publication, U.S. Govt. Printing Office, 1981).
- [14] Turner, J. S., Discussion Meeting, Kinetics of Physicochemical Oscillations, Aachen, September, 1979, and to be published.
- [15] Ruelle, D., private communication.
- [16] Packard, N. H., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S., *Phys. Rev. Lett.* 45 (1980) 712.
- [17] Roux, J.-C., Rossi, A., Bachelart, and Vidal, C., *Phys. Lett.* 77A (1980) 391; *Physica D* (in press).
- [18] Lanford, O. E., Strange attractors and turbulence, in: Swinney, H. L. and Gollub, J. P. (eds.), *Hydrodynamic Instabilities and the Transition to Turbulence*, Topics in Applied Physics Vol. 45 (Springer, Berlin, Heidelberg, New York, 1981), p. 7.
- [19] Showalter, K., Noyes, R. M., and Bar-Eli, K., *J. Chem. Phys.* 69 (1978) 2514.
- [20] Kac, M., private communication.
- [21] Field, R. J. and Noyes, R. M., *J. Chem. Phys.* 60 (1974) 1877.
- [22] Field, R. J., *J. Chem. Phys.* 63 (1975) 2289.
- [23] Feigenbaum, M. J., *J. Stat. Phys.* 19 (1978) 25.
- [24] Vidal, C., in: Haken, H. (ed.), *International Symposium in Synergetics* (Springer, Berlin, Heidelberg, New York, 1981).
- [25] Tomita, K. and Tsuda, I., *Prog. Theor. Phys.* 64 (1980) 1138.
- [26] Hudson, J. L. and Mankin, J. C., *J. Chem. Phys.* (June, 1981).

Nonlinear Plasma Dynamics Below the Cyclotron Frequency

John M. Greene
Princeton Plasma Physics Laboratory
Princeton, New Jersey, 08544

Plasma is a soup of electromagnetic fields, ions, and electrons. The essential aspect of this state of matter is that the charged particle components interact strongly with the electromagnetic field. There are a variety of plasma regimes. Here I am thinking of a hot, completely ionized gas so that there are no neutral atoms in the soup. This sort of plasma can be found in a number of places, such as thermonuclear fusion devices, the magnetosphere, and a variety of astrophysical locations.

A good way to know a plasma is through its waves. Each of the independent components that compose it has its own modes of oscillation, and these are generally coupled together. A full description of a plasma this way is extremely complicated. However, if one restricts oneself to the study of modes with frequencies below the ion cyclotron frequency the situation is considerably simplified. Since I have spent most of my career working in this frequency range, this talk will be limited appropriately. On the other hand, this talk would seem to be a good opportunity to speculate, so I will wander beyond the bounds of calculations that I have completed.

The electrons are very light, so their vibration frequencies are high. Thus in the low frequency regime they closely follow the ions. Furthermore, the ions all tend to move together over the course of these slow time scales. The appropriate model for these circumstances is one that treats the plasma as a single fluid. Only its mass density, momentum density, and stress need be considered. Further, since the light electrons react rapidly to screen out electric fields, only the magnetic part of the electromagnetic field need be treated completely.

These considerations lead on to the ideal MHD equations,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{v} = 0$$

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot [\rho \mathbf{v} \mathbf{v} + (p + B^2/2) \mathbf{I} - \mathbf{B} \mathbf{B}] = 0$$

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) = 0$$

$$\frac{\partial \rho S}{\partial t} + \nabla \cdot \rho S \mathbf{v} = 0$$

$$p = \rho^2 \frac{\partial}{\partial \rho} U(\rho, S)$$

The first of these is conservation of mass density, ρ . Next is the conservation of momentum, with \mathbf{v} the fluid velocity. The fluid stress has been taken to be a scalar, the pressure p , and \mathbf{I} is the unit tensor. The equation for the evolution of the magnetic field \mathbf{B} is the induction equation with an electric field that vanishes in the frame moving with the fluid velocity,

$$\mathbf{E} = - \mathbf{v} \times \mathbf{B}$$

Consistent with the neglect of viscosity in the momentum equation and resistivity in the induction equation, the energy equation has been reduced to the conservation of entropy, S . This is then related back to the pressure through the internal energy, U .

The first thing to notice about these equations is that the $\rho \mathbf{v} \mathbf{v}$ and $\mathbf{B} \mathbf{B}$ terms in the momentum flux have a similar form. As a result there are useful analogies between fluid dynamics, with $\mathbf{B}=0$, and ideal MHD. In another close analogy, within the ideal MHD model the magnetic flux and magnetic lines of force are convected with the fluid just as vorticity is convected in inviscid fluid flow.

The next thing to notice is that these equations are a generalization of the Euler equations of fluid dynamics. Analogously, one can expect that nonlinear flows develop singularities in finite time, leading to shocks. That is, the sources of entropy have been assumed to be sufficiently weak in relation to the wavelengths of the problem under investigation, whatever it may be, that dissipation is primarily confined to thin layers. Shocks are indeed a feature of this overview, but there are some complications arising from the plasma state and plasma waves that will be discussed below.

An idea of the waves existing in this model can best be obtained by studying an infinite homogeneous medium. The dispersion relation can be displayed schematically as in Fig. 1.

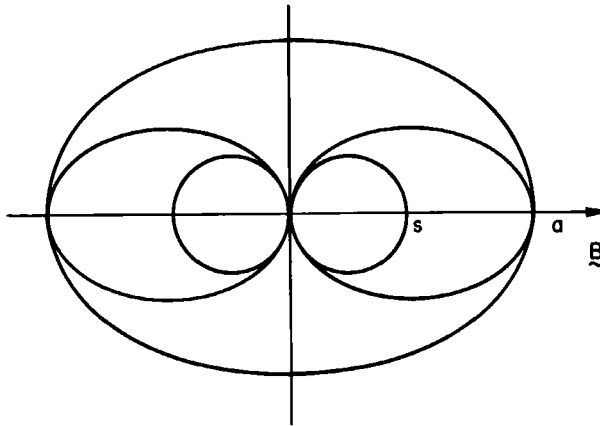


Fig. 1. The phase velocity of plane waves in a plasma as a function of the angle between the normal to the plane wave and the magnetic field direction. s is the sound speed and a is the speed of the Alfvén modes.

Figure 1 illustrates the phase velocities of plane waves as a function of the angle between the magnetic field and the normal to the constant phase surfaces, where the horizontal axis is the direction of \mathbf{B} . Since an evolution equation for the magnetic field is included in the set above, there are transverse as

well as longitudinal components to these waves. The coupling of the two transverse and one longitudinal polarization yield the three modes illustrated in Fig. 1. In this figure the characteristic speed of the longitudinal ("sound") modes, $s^2 = \partial p / \partial \rho$, has been taken to be smaller than the characteristic speed of the transverse (Alfven modes, $a^2 = B^2 / \rho$). The outer curve describes a wave that spreads more or less uniformly in all directions, in the manner of an ordinary sound wave, but with approximately the Alfven speed. It is known as the fast magnetosonic mode. The middle curve is purely transverse, and will be called the shear Alfven mode. The inner curve is the slow magnetosonic branch of the dispersion relation. The important feature of this sketch for the remainder of this review is the vanishing of the phase speed for the two slowest modes propagating perpendicular to the magnetic field.

To obtain a deeper understanding of these equations, it is useful to consider a specific configuration. The primary example used here will be a tokamak. This is an axisymmetric configuration with magnetic field lines winding around toroidal surfaces, and with a hot plasma confined to the inner toroids. A cross section of the magnetic surfaces that contain the field lines can be sketched as in Fig. 2.

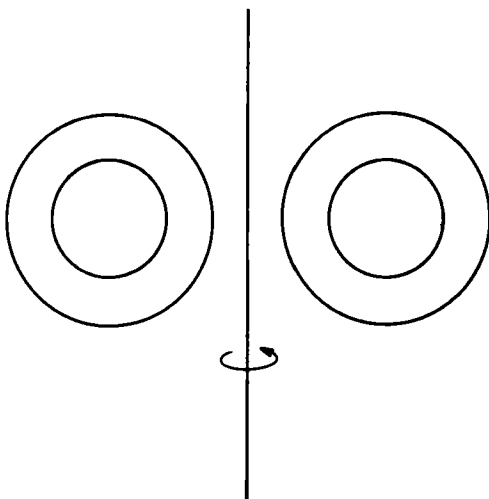


Fig. 2. A cross section of magnetic field lines in a tokamak.

Our method of applying the MHD equations to this configuration has been to first calculate an equilibrium, next to treat its linear stability, and finally to consider the nonlinear meaning of any linear instability. Much of the effort has gone into the linear problem. Incidentally, the configurations we call, "equilibrium" are equivalent to the configurations that Orszag called "quasi-equilibrium" in another paper in this collection [1]. They are in equilibrium with respect to the generation of waves through the Euler equations, but may be changing slowly due to the diffusion of mass, momentum, energy, and magnetic flux. Our equilibria are two-dimensional as were those of Orszag.

The linear stability problem is characterized by singularities that are very similar to those of the inviscid limit of the Orr-Sommerfeld equation. One of the first people to utilize these singularities in a stability calculation was Suydam at Los Alamos [2]. The singularities of the inviscid Orr-Sommerfeld system arise where the fluid streaming velocity is equal to the phase velocity of the perturbation. That is also the case here, with the variation that this velocity vanishes. It occurs along closed field lines for perturbations that

have constant phase along these lines. Thus these perturbations represent waves propagating perpendicular to the magnetic field at that point. Remember that two modes have vanishing phase velocity in this direction. Since these field lines wind around the toroidal magnetic surfaces, the singular perturbations are three-dimensional.

Closer examination of these singularities from a nonlinear viewpoint shows that they arise from an attempt to change the topology of the magnetic surfaces, as illustrated in Fig. 3.

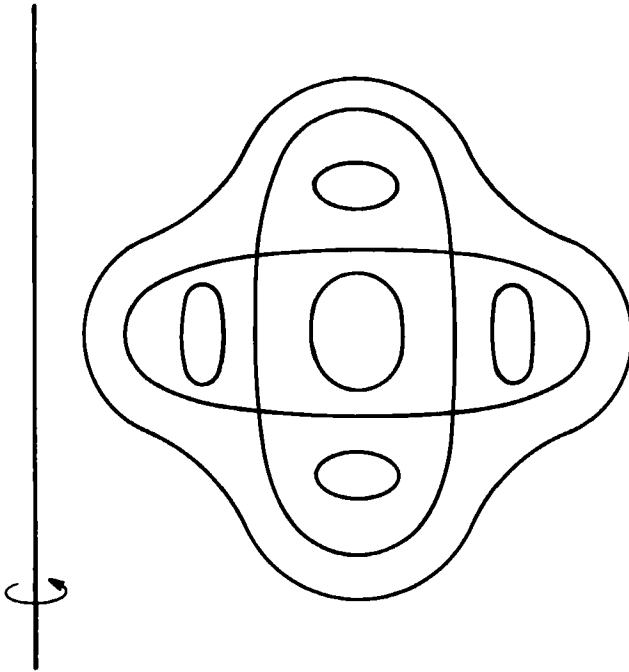


Fig. 3. A schematic sketch of cross sections of surfaces containing magnetic field lines showing "island structure" and "x-points."

Again, these lines are the cross section of surfaces that contain the magnetic field lines. What will be called an island structure has opened out of a toroidal surface of intermediate radius. Its associated tube of flux winds helically around the tokamak. Thus the line of force at the center of the flux cell, and also the line at the junction of the cells, close after traversing the tokamak four times, for the case shown here. This change in topology is inconsistent with convection of magnetic field lines with the fluid. The strings of fluid elements connected by field lines are rearranged as the magnetic flux and fluid flows into and fattens out the fourfold island structure illustrated in the sketch above. There is reconnection of these strings at the points where the islands join. Since these points are intersections of the magnetic surfaces that separate the different flux regions, they are called x-points. Thus an essential aspect of nonlinear plasma dynamics below the cyclotron frequency is the study of reconnection at x-points.

The first problem that arises when studying x-points is to describe what is meant by the term. The first step in much of x-point theory is to draw an x on the blackboard (or the page, in the present case), as in Fig. 4.

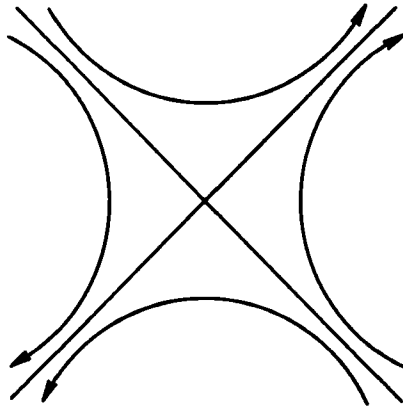


Fig. 4. A schematic illustration of magnetic field lines near an x-point.

and study problems with a translational invariance perpendicular to the page. This is a useful approximation in many respects, but it does not contribute to the identification of x-points in complicated configurations. This page representation of an x-point treats a plane orthogonal to a chosen line of force. The magnetic field in this plane vanishes at the x-point, since that is how the representation was chosen. However, there is no reason for the total magnetic field to vanish there. Indeed, if the magnetic field vanishes along a line in some configuration, almost any small perturbation will yield a finite magnetic field everywhere except at a discrete set of points.

The conclusion from this line of thought is that there is no way to identify an x-point from local considerations. An x-point is the consequence of global relations.

Returning to the tokamak configuration, the singularity that produced the x-point is associated with a closed field line. In Fig. 3 the x-points are the closed field lines that lie at the juncture of different flux cells. This is the general description of a tokamak x-point.

Cowley [3] has discussed configurations in which reconnection and x-point behavior occurs without a closed field line of the tokamak type. From his work it is clear that there is another kind of x-point. This configuration has not been thoroughly exploited, even though it is probably very common in the magnetosphere, the sun, and elsewhere. Briefly, these x-points are the magnetic lines that connect points at which the magnetic field vanishes. This will be discussed in more detail in the next few paragraphs.

If the magnetic field points in every direction in the neighborhood of a point at which the field vanishes, a perturbing field can move the null point, but cannot destroy it. Consider the local behavior around the null point. If the coordinates are chosen so that this point lies at the origin, the local magnetic can be given in terms of a constant matrix, δB ,

$$\vec{B} = \delta B \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

The condition that \mathbf{B} be divergence free becomes

$$\text{Tr } (\delta\mathbf{B}) = 0$$

From this it follows that two eigenvalues of $\delta\mathbf{B}$ have one sign and the third has the other sign. Thus the magnetic nulls can be divided into two classes, those for which the odd eigenvalue is positive, and those for which it is negative. Following Cowley, we will call the first class type A nulls, and the second class type B.

Next consider the lines that asymptotically approach or diverge from the null point. The eigenvectors of $\delta\mathbf{B}$ define three such lines. Further, consider the plane defined by the two eigenvectors whose corresponding eigenvalues have the same sign. Then all magnetic lines that asymptotically lie in this plane either approach type A null points or diverge from type B null points. Extending this plane by following lines of force then defines a two-dimensional manifold of lines that are asymptotic to a given null. We will call this manifold the sheet associated with the given null.

Now consider the intersection of two sheets. Since each sheet is composed of lines of force, the magnetic field cannot have a component perpendicular to either sheet. Thus the line of intersection of two sheets must be a magnetic field line. Since each line in one sheet must diverge from a B null, and each line in the other sheet must converge toward an A null, this line of intersection must be a null line.

A sketch of the near intersection of two sheets, such as this,

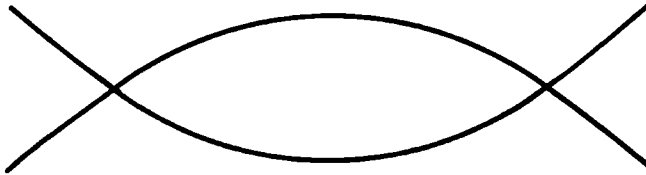


Fig. 5. A sketch of the near intersection of two sheets, illustrating that null to null lines are formed in pairs.

can easily convince one that null to null lines are created or destroyed in pairs.

To show that null to null lines lie at the juncture of different flux regions, and thus qualify as x -points, requires some complex three-dimensional visualization. Cowley worked out a particular example, a uniform field orthogonal to the axis of a point dipole. They can be called the solar wind and earth fields, respectively. A sketch of a cross section of this configuration through the null points is shown in Fig. 6.

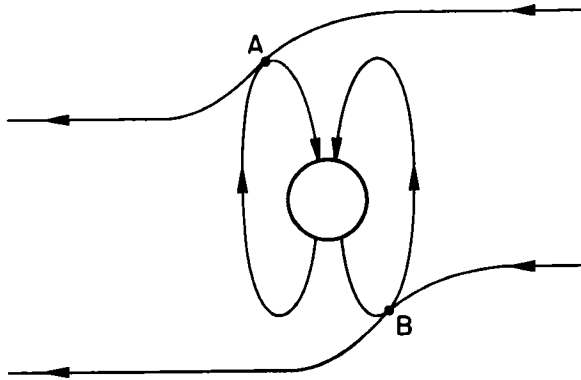


Fig. 6. A schematic illustration of the null lines resulting uniform field orthogonal to the axis of a point dipole.

The null to null lines lie on a circle whose plane is orthogonal to the paper. Clearly, the sheet of lines that go into the null point A cannot decide whether to dive into the earth, or wander out into space, and thus wind up at A. Those that diverge from B have the same problem with the opposite orientation. Thus these sheets divide flux bundles that are defined by the asymptotic behavior of the field lines that compose the bundle. The null to null lines are common to both sheets, and thus are at the juncture of all the flux bundles. In this sense, they should be quite acceptable x-points.

The program of investigating x-point behavior on null to null lines remains to be carried out in detail. No doubt a thorough investigation would modify the picture suggested here. Still, the idea that null to null lines are x-points would seem to be based on solid intuition. The events that happen when two null to null lines are thrust into existence by the intersection of two sheets might be particularly interesting.

A central problem of x-point theory is to explain how fast the islands of Fig. 3 can fatten up. To what extent does the x-point configuration impose narrow dissipation layers or shocks that promote rapid breaking of the topological constraints of the Euler equations? This is a large and controversial subject. It is impossible to do justice to it here, but it cannot be ignored in this overview either. A fairly recent review paper has been published by Vasyliunas [4].

A central feature of an x-point that is related to this question is the angle between the two bars that compose the x. The component of \vec{B} in the plane of the page can be written in terms of a stream function,

$$\vec{B} = e_z \times \nabla \psi$$

The stream function can be decomposed into a part that is imposed from outside, and part that is due to current flowing at the x-point,

$$\psi = I^0(x^2 - y^2) + j(x^2 + y^2)$$

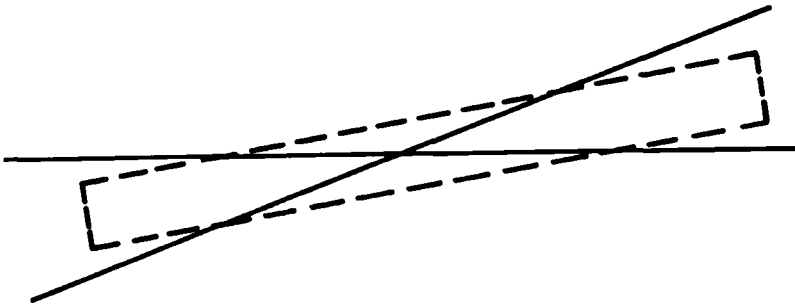


Fig. 7. A schematic illustration of the tendency to form a current sheet of an x-point.

where I^0 is the magnitude of the imposed quadrupole field and j is the current density at the x-point. Thus if the x-point sits in a vacuum field the bars meet at right angles. Larger current densities can be accommodated at an x-point if they lie on a bar, so that the outer parts contribute to the imposed quadrupole field at the x-point, Fig. 7. Thus any movement of fluid toward and through an x-point, carrying magnetic field lines with it, is going to act to close the angle of the x, and create a current sheet. Consequently, there will be large forces on the fluid in this region. Here we have a region of nonlinear flow. It seems reasonable that there should be shock waves around. This was suggested by Petschek [5] in a famous paper some years ago. He drew a picture as in Fig. 8, where the dotted lines denote shocks.

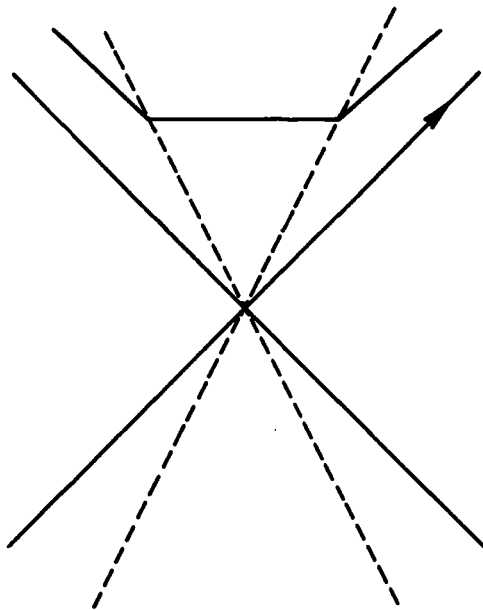


Fig. 8. A schematic illustration of the shock wave fronts formed near an x-point.

The best recent work on this configuration that I know of has been done by Hamieri [6]. I am not going to get into this, but only deal with one slightly tangential aspect.

Dealing with a fluid like air, one is comfortable considering a shock as a line. Shock widths are of the order of a mean-free path, which is smaller than any length of interest. For problems dealing with flow through x-points, however, that is not so clear. The shock structure may engulf and transform a significant volume, particularly near the x-point where several shocks converge. Similar problems have arisen in plasma physics before. Crude estimates of the thickness of shocks of the fast magnetosonic wave mode yielded unreasonable shock dimensions. Much theoretical and experimental effort has been invested in a thorough study of the actual shock structure associated with this mode. This resulted in the discovery of the "collisionless fast magnetosonic shock."

A similar effort applied to the slow magnetosonic mode might produce very interesting results. Slow magnetosonic shocks are no doubt very different from fast magnetosonic shocks, since the waves themselves are very different. A major difficulty with the Petschek mechanism for flow through an x-point is that we do not know with certainty what is meant by the shock lines drawn in this picture.

There is one other aspect of island formation in a tokamak that will be treated in this overview. That is the question of wandering field lines. In drawing Fig. 3, it was assumed that magnetic field lines lay on surfaces. That is not always true. There are configurations in which a field line, on successive interactions of the cross section shown in Fig. 3, will wander erratically over the page. This is bad news for tokamaks. Fluid and heat flow more or less unimpeded along field lines, so a tokamak with such a configuration is not a good containment device.

Following a magnetic field line around a tokamak from one intersection of a cross section to another defines a mapping of the cross section onto itself. This map conserves flux, since a given area is mapped into another with the same flux. Conservation of flux differs only trivially from the conservation of area, and thus is equivalent to Poincare's invariant for Hamiltonian systems. Hence the tracing of field line trajectories is closely related to a wide variety of mechanical and celestial problems, and results from one field can be applied to another.

As a result, one can make rather general statements about the behavior of magnetic field lines that are true of nearly all configurations subject to the constraint that $\nabla \cdot \mathbf{B} = 0$. Consider, for example, an ideal tokamak as illustrated in Fig. 2. Here the field lines loop around toroidal surfaces with a winding number that varies continuously from surface to surface. When such a configuration is perturbed with an arbitrary magnetic field, each surface on which the winding number is rational sprouts an island structure such as illustrated in Fig. 3. Thus there are dense sets of islands of all multiplicities. However, if the magnetic field is further restricted by the condition that the magnetic stress of Eq. (1) be reasonable, these islands do not appear. That is, modes that produce islands of high multiplicities are stable according to MHD theory. Of course, there is still the possibility of excitation of these modes by thermal fluctuations, and that may be a very interesting topic of investigation, but basically the tokamak configuration appears to be preferred experimentally.

The exception to this rule are islands with low multiplicities, two, three, or four, as shown in Fig. 3. These observed and can grow to quite large amplitude. In the remainder of this paper I will speculate on their effect on the stability of very nearby high multiplicity islands.

Since the winding numbers of the magnetic field lines in the ideal tokamak are continuous, the field line that closes after 25 rotations around the cross section shown in Fig. 3 while making 101 loops is highly contorted by the growth of the four-fold island. This should have a considerable effect on its stability properties. That is, the self-consistent constraint that the magnetic stress be acceptable is probably less constricting in this circumstance. Thus, there should be a 101-fold island and all others with a nearby winding number, clustered around the fourfold island.

At the center of each of these islands is a closed field line. The island flux tube winds around this line, something in the nature of a small, thin, helical tokamak. The rate of rotation within this flux tube is a quantity of interest. As increasing flux and current are drawn into the island this subsidiary rotation increases. A critical point is reached when the center of the flux tube rotates at half the rate at which the center line winds around the tokamak. Then, in almost every case, the center of the flux tube takes the form shown in Fig. 9.

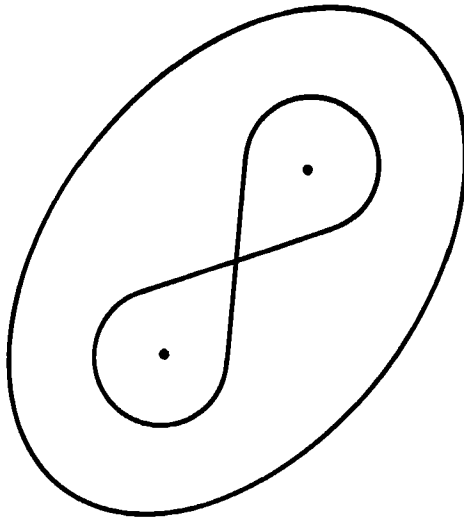


Fig. 9. Illustration of the formation of an x-point, and subsequent period doubling bifurcation, in an island flux tube in a tokamak.

That is, the central magnetic line becomes an x-point, and a new central line of twice the length is created. This is known as a period doubling bifurcation.

Following this argument on, if the driving fourfold island further increases in size, more fluid and flux will be forced into the 101-fold island, and into the interior 202-fold island. It too can undergo a period doubling bifurcation.

As one might expect here at Los Alamos, in general systems, without the magnetic stress constraint, an infinite sequence of doubling bifurcations are produced by a finite driving perturbation, and these bifurcations have a universal behavior. It remains to be seen if the stress constraint effects this universality. I would guess that it won't.

At each stage of the bifurcation new x-points, lines from which all other lines diverge, are created. In the limit of infinite bifurcations, then, all lines diverge from each other. This is the essential character of wandering field lines. Thus as the fourfold island grows it will be surrounded by an expanding sea of wandering field lines. When this sea laps clear out to the edge of the tokamak containment region, there is a catastrophe. Certainly the sudden spreading of contained plasma in a "disruptive instability" is a major constraint on tokamak operation. If my interpretation is correct, and the observed disruptive spreading is due to sudden onset of wandering fields, universal period doubling bifurcations are an important feature of tokamaks.

REFERENCES

1. A. T. Patera and S. A. Orszag, this volume.
2. B. R. Suydam, Proc. 2nd Int. Conf. Peaceful Uses of Atomic Energy, Geneva, Vol. 31, p. 157 (1950).
3. S. W. H. Cowley, Radio Science 8, 903 (1973).
4. V. M. Vasyliunas, Rev. Geophys. Space Phys. 13, 303 (1975).
5. H. E. Petschek, NASA Spec. Publ. SP 50, 425 (1964).
6. E. Hamieri, J. Plasma Phys. 22, 245 (1979).

This Page Intentionally Left Blank

Turbulence and Self-Consistent Fields in Plasmas

D. Pesme and D. DuBois

Theoretical Division
University of California
Los Alamos National Laboratory
Los Alamos, NM 87545 USA

This paper is concerned with the role of self-consistency of the electric field in 1-D plasma turbulence. We first show that in the non-self consistent electric field problem excellent agreement is found between numerical experiments and quasilinear theory whenever the imposed electric field Fourier components have random phase. A discrepancy is exhibited between quasilinear prediction and numerical simulations in the self-consistent electric field case. This discrepancy is explained by the creation of a long correlation time of the electric field resulting from a strong wave-particle interaction. A comparison is made between quasilinear and renormalized propagator theories, and the Dupree "Clump" theory. These three theories are found to be self-contradictory in the regime of strong wave-particle interaction because they make an a priori quasi-gaussian assumption for the electric field. The Direct Interaction Approximation is then shown to be a closure taking into account in a self-consistent way non-gaussian effects.

INTRODUCTION

This paper is concerned with the role of self-consistent electromagnetic fields in the theory of turbulent plasmas. In recent years dramatic progress has occurred in our understanding of the motion of charged particles in prescribed electromagnetic fields: We have learned that there are thresholds for the onset of chaotic motion which are governed by conditions such as resonance (or island) overlap; the chaotic motion is characterized by diffusion of particles in velocity and by the exponential divergence of initially nearby trajectories. Above the stochastic threshold particles can wander all over and indeed their trajectories appear to completely fill certain regions of phase space. The detailed studies of intrinsic stochasticity have concentrated on the prescribed field problem.

In this paper we will try to make a preliminary assessment of the importance of the self-consistency of the electric fields in the turbulent or chaotic behavior of plasmas. We will first review in Section 2 some numerical studies of prescribed field models in parameter regimes well above the stochastic threshold. In certain cases it is known that, even for a single realization of the imposed field, a statistical theory on the level of the familiar quasilinear theory provides an accurate description of the diffusion in velocity space when compared to numerical solutions. It is less well-known that the exponential separation of orbits is likewise accurately predicted by a statistical theory of the two-point correlation function. In Section 3 we review the results of statistical theories which involve the ensemble average over many realizations of the imposed fields; in particular the imposed electrical fields are considered to be quasi-

gaussian random variables, in a sense which will be defined in Section 2. Usually this property is realized by assuming random phases for each mode. We examine the conditions under which the predictions of such a statistical theory are able to describe the properties of a single realization such as a numerical simulation run.

In Section 4 we will review some aspects of the self-consistent field case. Self-consistent numerical simulations in the quasilinear parameter regime are not in agreement with the imposed field statistical theories. The reason is the following: the self-consistent Maxwell equations are nonlinear and impose phase correlations between modes. This invalidates the random phase or quasigaussian fluctuation assumption and introduces new correlation times into the problem; this is shown in an iterative calculation in Section 5. The imposed field statistical theories cannot be made self-consistent a posteriori because they neglect important nongaussian correlations.

In a sense (to be discussed) the assumption that E is a gaussian random variable is an assumption of maximum chaos. Since self-consistency invalidates this assumption then a self-consistent system must have less than maximum chaos. Self-consistency implies a kind of self-organization of the chaos.

The formulation of a self-consistent, statistical description of a turbulent plasma is being pursued with renewed vigor at the present time. An approach that shows considerable promise involves the application of the strong turbulence approximation known as the direct interaction approximation (DIA) to Vlasov turbulence. This approximation incorporates certain nongaussian correlations into a statistical theory that retains the basic conservation laws and other realizability properties. In Section 5 we will examine certain features of this theory and make some comparisons with the imposed field theories.

REVIEW OF INTRINSIC STOCHASTICITY STUDIES IN PRESCRIBED FIELDS

The imposed field studies seek to establish the properties of single particle trajectories in prescribed electric and magnetic fields. Here we will restrict our considerations to one-dimensional electrostatic models with electric fields of the form

$$E(x, t) = \sum_{n=1}^{N_m} \sum_k |E_k| \exp(i\theta_k) \exp[ikx - \omega_n(k)t] \quad (2.1)$$

We consider a finite system of length L so that the k modes are discrete and we allow the possibility of a set of frequencies $\omega_n(k)$ associated with each mode, where $1 \leq n \leq N_m$. For simplicity we will write most of our equations explicitly for the case $N_m = 1$ and $\omega_1(k) = \omega(k)$. Newton's equations for the motion of a particle of charge q and mass m in this field are of course:

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= \frac{q}{m} E(x, t) = \frac{q}{m} \sum_k |E_k| \exp[ikx - \omega(k)t + \theta_k] \end{aligned} \quad (2.2)$$

and are highly nonlinear. We first focus on a single resonance associated with a given mode k_0 ; Newton's equation for a particle interacting with this single mode can be written

$$\dot{x} = v \quad (2.3)$$

$$\dot{v} = \frac{g}{m} E_{k_0} \sin[k_0 x - \omega(k_0)t]$$

It is convenient to transform to the coordinates X, V in the wave frame:

$$x = v_\phi(k_0)t + X \quad (2.4)$$

$$v = v_\phi(k_0) + V$$

where $v_\phi(k_0) = \omega(k_0)/k_0$ is the wave phase-velocity. In this frame the equations of motion are

$$\begin{aligned} \dot{X} &= V \\ \dot{V} &= \frac{g}{m} E_{k_0} \sin k_0 X \end{aligned} \quad (2.5)$$

These are the well-known equations for the nonlinear pendulum whose solution is exactly known. The phase space is divided into trapping and passing regions as shown in Fig. 1 by the separatrix whose equation is

$$V = \pm \left[2 \frac{g}{m} \frac{|E_{k_0}|}{k_0} \right]^{1/2} \cos \frac{k_0 X}{2} \quad (2.6)$$

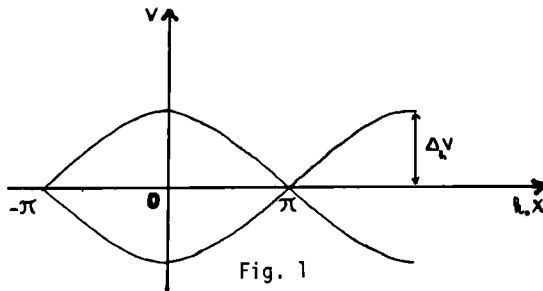


Fig. 1. Separatrix in phase-space for a single resonance.

The "trapping half width" of the islands of the separatrix is

$$\Delta v_h(k_0) = \sqrt{2 \frac{g}{m} \frac{|E_{k_0}|}{k_0}} \quad (2.7)$$

In the case of many modes we can label each resonance by its phase velocity and trapping half-width as if it were an isolated resonance. The distance in v between the phase velocities of the isolated resonances is

$$\begin{aligned} \delta v &= v_\phi(k_{n+1}) - v_\phi(k_n) \\ &= \frac{\partial}{\partial k} v_\phi(k) \times \frac{2\pi}{L} \end{aligned} \quad (2.8)$$

where $k_n = 2\pi n/L$. For very small amplitudes the resonances, at a first approximation, can be considered as isolated and we have the picture of superimposed resonances shown in Fig. 2.

Fig. 2. Approximate resonance domains in phase space for several resonances such that $k_s \ll 1$.

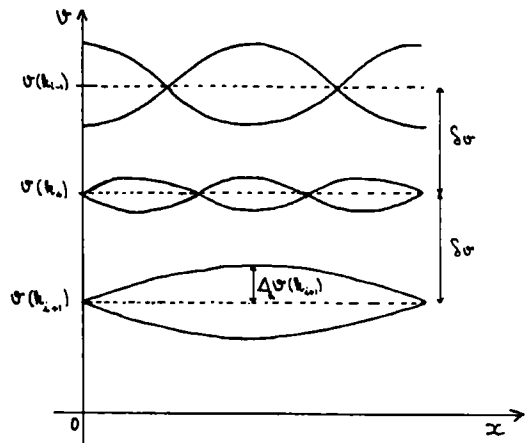


Fig. 2

The work of Chirikov¹ showed that it was useful to define the stochasticity parameter k_s :

$$k_s = \frac{\Delta V_h(k_n) + \Delta V_h(k_{n+1})}{\delta v} \quad (2.9)$$

He showed that for $k_s \ll 1$ the motion is regular (except in small disconnected regions of phase space) but for $k_s > 1$ it is chaotic in the phase space domain for which the velocity belongs to the interval of phase velocities of the resonances. As an example we display the results of Escande and Doveil² for the case of two well-separated resonances with the equations of motion

$$\dot{x} = v$$

$$\dot{v} = M_1 \sin x + M_2 \sin k(x-t)$$

where $v_\phi(k) = 1$ and $v_\phi(0) = 0$. They computed the orbits $x(t)$, $v(t)$ as a function of t from some initial conditions x_0 , v_0 at $t = 0$. The results are summarized in Fig. 3 for the case $k_s \ll 1$. The primary resonances are distorted as evidenced by the deviation of the separatrices from the sinusoidal shape. In addition, new secondary resonances appear. The trajectories near the separatrices become fuzzy forming a stochastic layer. The important result is that there still exists definite passing trajectories and these isolate the phase space in the sense that a particle cannot travel from one set of trapping islands to another. Thus the motion is stable and $|v(t) - v_0|$ is bounded so a particle can't wander far in velocity space; thus there is no diffusion. As k_s is increased (but $k_s \lesssim 1$) the width of the stochastic layer grows as does the size of the

secondary resonances. There exists a value of k_s (near unity) where all the stochastic layers join together at which point the system undergoes a gross stochastic instability in which particles can wander over the phase space between the resonances. Chirikov originally postulated that the stochastic threshold occurred at $k_s = 1$. More recent work^{3,2} has shown that the precise value of the threshold is somewhat less the unity. In the following we will assume $k_s \gg 1$, well above the chaotic threshold. We will not attempt a precise definition of chaos here; suffice it to say for our purposes, the chaotic regime is characterized by diffusion of particles in velocity and exponential divergence of nearby orbits. The former is a very practical consequence of chaos the latter is a more theoretical consequence related to the rate at which information is lost by the system as characterized by the so-called Kolmogorov entropy.

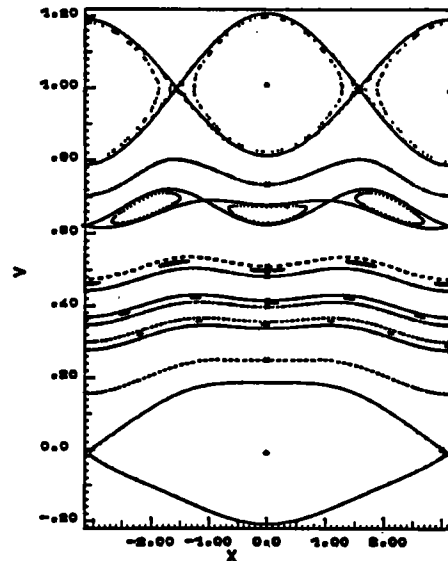


Fig. 3. Distortion of primary resonances and appearance of secondary resonances for $k_s = 0.4$ (from (Escande et al., Ref. 2).

Fig. 3

There have been many numerical studies of these quantities for various models of the imposed field of the form in equation (2.2) which appear to fall into three classes:

- i) The first class considers only a single wavenumber but associates an infinite number of frequencies (harmonics) with this wave number (i.e., $\omega_n(k) = n\omega_0$ and $N_m = \infty$). This class includes the famous Chirikov-Taylor mapping and has been investigated by Rechester and White⁴ and others.⁵
- ii) A second class of models has an infinite (or large) number of modes k each with an infinite number of harmonics of a fundamental frequency (i.e., $\omega_n(k) = n\omega(k)$ and $N_m = \infty$).

Such models have been investigated by Rechester, Rosenbluth and White,^{6,7} Molvig et al.⁸

- iii) The class of models closest to our interests have an infinite (or large) spectrum of k values each with a definite frequency $\omega(k)$ (i.e., $N_m =$

1 and $\omega_1(k) = \omega(k)$). Simulations of this case have been carried out by Flynn⁹ and Doveil.¹⁰ In these numerical calculations values of $|E_k|$ for Eq. 2.2 were assigned and the phases θ_k were randomly chosen between 0 and 2π with no correlation between different values of k . This is usually called the Random Phase Approximation (RPA). This procedure generates a stochastic process which is asymptotically equivalent to a gaussian process in the limit of a large number of modes.¹¹ S. P. Gary and D. Montgomery,¹² K. N. Graham and J. A. Fejer¹³ also did similar calculations, though they used somewhat different processes than random phases in order to generate a quasigaussian electric field. In the following, for simplicity, we will consider in the imposed electric field case only the models with random phase Fourier components. The single particle trajectories $x(x_0, v_0, t)$ and $v(x_0, v_0, t)$ were computed numerically from initial conditions x_0, v_0 at $t = 0$. Velocity diffusion was studied by computing the spatial average over the initial positions of the squared velocity deviation

$$\overline{\Delta v^2(t)} \equiv \overline{(v(x_0, v_0, t) - v_0)^2}^{x_0} \quad (2.10)$$

where we define the spatial average of any function $\psi(x)$ by

$$\overline{\psi}^{x_0} \equiv \frac{1}{L} \int_0^L \psi(x_0) dx_0$$

Results from these numerical calculations showed that:

i) Δv^2 was proportional to t (except for very short times). The observed velocity diffusion coefficient in the numerical experiments can thus be defined as

$$D_{\text{exp}} (\text{def}) \equiv \frac{1}{2t} \overline{\Delta v^2(t)} \quad (2.11)$$

ii) The observed diffusion coefficient was in good agreement with the prediction of quasilinear theory

$$D_{\text{exp}} = D^{Q.L.} \quad (2.12)$$

where

$$D^{Q.L.} = \frac{q^2}{m} \sum_k^2 \sum_n |E_k|^2 \int_0^t dt \exp i [\omega_n(k) - kv] \tau \quad (2.13)$$

The quasilinear theory is a statistical theory, averaged over many realizations of the phases, θ_k , which are assumed to be random. We will discuss statistical theories in the next section.

Similar good agreement between the observed diffusion coefficient and the predictions of quasilinear theory was found in the work of Rechester, Rosenbluth and White⁷ as it can be seen in Fig. 4. For their model the quasilinear prediction is $D^{Q.L.} = \epsilon^2 M/4$ where ϵ is the amplitude of the Fourier components and M is the number of harmonics (Nm in our notation).

Fig. 4. Comparison between numerical experiments (dots ϕ) and the quasilinear prediction (solid line) for the value of the diffusion coefficient (from Rechester et al., Ref. 2).

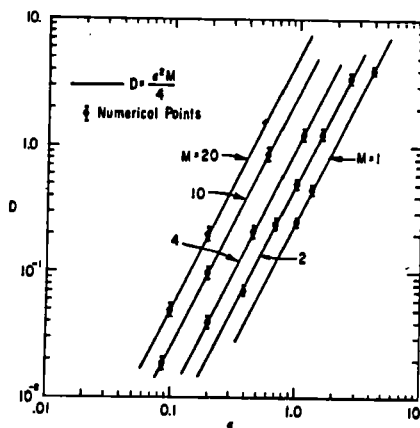


Fig. 4

Rechester, Rosenbluth and White also computed the so-called "Kolmogorov entropy" defined as

$$h = \lim_{t \rightarrow \infty} \frac{1}{t} \lim_{d_0 \rightarrow 0} \ln \left[\frac{d(t)}{d_0} \right] \quad (2.14)$$

where d is the "distance" between two nearby orbits. We can take this to be

$$d(t) = \left[k_+^2 (x_1 - x_2)^2 + \frac{(v_1 - v_2)^2}{v_+^2} \right]^{1/2} \quad (2.15)$$

where $x_1 = x(t, x_{01}, v_{01})$, $v_1 = v(t, x_{01}, v_{01})$ and $x_2 = x(t, x_{02}, v_{02})$, $v_2 = v(t, x_{02}, v_{02})$ corresponding to different but nearby initial conditions at $t = 0$; $d_0 = d(t = 0)$. The normalization constants in this expression are $v_+ = \frac{1}{2}(v_1 + v_2)$ and $k_+ = k(v_+)$; $k(v)$ is the solution of the implicit equation $\omega(k(v)) = k(v)v$. Positive values of h are interpreted to be a signature of chaotic behavior. Roughly speaking, in physical terms, this definition implies that for short times $d(t) \simeq d_0 \exp ht$ and this implies a loss of memory of initial conditions and a sufficient condition for mixing and ergodicity.

The results of the numerical experiments of Rechester et al.,^{6,7} are shown in figure 5. They showed that h_{exp} exists and is independent of initial conditions if d_0 is sufficiently small and obtained excellent agreement with the prediction $h^{Q.L.}$ obtained from a statistical theory assuming a quasigaussian electric field. This theory, which is discussed in the next section provides an explicit relation of $h^{Q.L.}$ to the quasilinear diffusion coefficient

$$h^{Q.L.} = 0.36 \left(k_+^2 D^{Q.L.} \right)^{1/3} \quad (2.16)$$

Fig. 5. Comparison between numerical experiments (dots ϕ) and the quasilinear prediction (solid line) for the value of the Kolmogorov entropy (from Rechester et al., Ref. 6, 7).

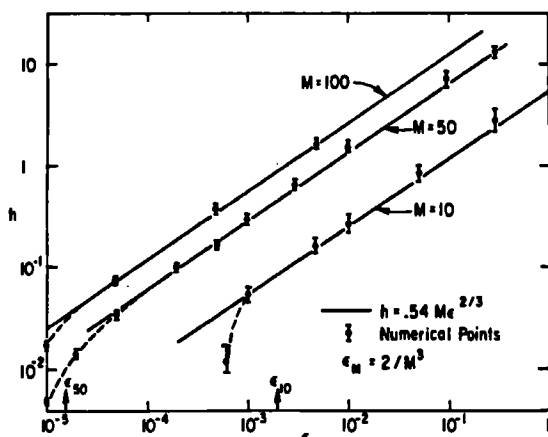


Fig. 5

The numerical experiments showed that to excellent accuracy

$$h_{\text{exp}} = h^{\text{Q.L.}} = 0.36 \left(k_{\text{+}}^2 \text{Q.L.} \right)^{1/3} \quad (2.17)$$

STATISTICAL TURBULENCE THEORIES FOR PRESCRIBED RANDOM FIELDS

We next generalize our considerations to a system of many-charged particles moving in electrostatic fields. If the fields are prescribed, we have a system of independent particles whose trajectories have the properties discussed in Section 2. The many particle situation is conveniently described by the one particle phase space distribution function $f(x, v, t)$ which evolves after a time t from some initial distribution $f_0(x, v)$ at $t = 0$ according to the formula

$$f(x, v, t) = \int dx_0 \int dv_0 \delta[x - x(t, x_0, v_0)] \delta[v - v(t, x_0, v_0)] f_0(x, v) \quad (3.1)$$

where $x(t, x_0, v_0)$ and $v(t, x_0, v_0)$ are just the position and velocity discussed in Section 2. The phase-space density evolves according to the Liouville equation:

$$\left[\partial_t + v \partial_x + \frac{q}{m} E(x, t) \partial_v \right] f(x, v, t) = 0 \quad (3.2)$$

which expresses the conservation of density along phase-space trajectories. In the self-consistent case we can still use Eq. 3.2 in which $E(x, t)$ is the self-consistent field determined by Poisson's equation 4.1.¹⁴

In the prescribed field problem we assume $E(x, t)$ is a random variable over an ensemble each realization of which is given by an expression of the form (2.1). This is the so-called stochastic acceleration problem based on the linear stochastic equation (3.2). This problem has received much attention.¹⁵ We give here the results of this analysis in the quasilinear regime which involves the following assumptions

i) $k_s \gg 1$

ii) $h^{-1} \equiv \tau_h \ll \tau_{<f>}$ (3.3)

- iii) quasigaussian approximation for electric field fluctuations, e.g., phase θ_k random and uncorrelated.

Condition i) simply assumes that the system is well above the stochastic threshold. In condition iii), the quasigaussian approximation is defined as the case in which the correlation functions of any order are characterized by a unique correlation length. More precisely, throughout all this article, we name quasigaussian any random process for which an irreducible correlation function of any order is not zero only if all the arguments lie inside a same interval of length ℓ_c , where ℓ_c is the correlation length of the two point correlation function.^C This condition can be realized by the random phase approximation.

In condition ii) we have defined a characteristic time τ_h as the inverse of the Kolmogorov entropy; τ_h is a measure of the time over which information of initial conditions is lost. This time, τ_h , must be much less than the time of evolution of the mean distribution function, $\tau_{<f>}$. We can estimate the latter as

$$\tau_{<f>} \approx \frac{(\Delta v_f)^2}{D^{Q.L.}} \quad (3.4)$$

where Δv_f is a measure of the width in velocity space for which waves interact resonantly with particles and $D^{Q.L.}$ the quasilinear diffusion coefficient of Eq. 2.13. Note that $D^{Q.L.} \approx (q/m)^2 \langle E^2 \rangle \tau_c$ where $\tau_c = [\Delta k(v - v_g)]^{-1}$ is the correlation time of the electric field fluctuations as seen by a typical resonant particle; v_g is the group velocity of the wave with wave number $k(v)$. We can then write

$$\tau_{<f>}/\tau_c = (\omega_b \tau_c)^{-4} \text{ where } \omega_b = \frac{q}{m} \bar{k} \langle E^2 \rangle^{1/2}$$

is often called the effective bounce frequency; i.e., the frequency of a particle trapped in a monochromatic wave of amplitude $\langle E^2 \rangle^{1/2}$ and wave number \bar{k} . With these definitions it is found that condition ii) can also be written as

$$(\omega_b \tau_c)^4 \ll 1 \quad (3.5)$$

Under these conditions it is known¹⁶ that the equation of evolution of the ensemble averaged or mean distribution $\langle f \rangle$ is

$$\partial_t \langle f \rangle = \partial_v D^{Q.L.}(v) \partial_v \langle f \rangle \quad (3.6)$$

where $D^{Q.L.}$ is given in 2.13. The ensemble average is over many realizations of the electric fields. We assume statistical homogeneity, i.e. $\langle E(x, t) \rangle = 0$ and $f_0(x, v) = f_0(v)$.

An equation for the two-phase point, one time correlation function

$$\langle \delta f(x, v, t) \delta f(x', v'; t') \rangle, \quad \delta f(x, v, t) = f - \langle f \rangle,$$

can also be derived in this limit (Dupree¹⁷). In the limit of small separation in velocity space $|k(v_+)v_-| \ll \sup [(k^2 D^{Q.L.})^{1/3}, \gamma_k]$, and $|x_-/v_+| \ll \inf [(k^2(v_+)D^{Q.L.})^{-1/3}, \gamma_k^{-1}]$, this equation can be written as:

$$\left[\partial_t + v_- \partial_{x_-} - \left(D_{12}^{Q.L.} - D_{12}(x_-) \right) \partial_{v_-}^2 \right] < \delta f(x, v, t) \delta f(x', v', t) > \\ = 2D_{12}(x_-) [\partial_{v_+} < f(v_+, t) >]^2 \quad (3.7)$$

where D_{12} is the relative diffusion coefficient

$$D_{12}(x_-) = \left(\frac{q}{m} \right)^2 \sum_k \sum_n |E_k|^2 \text{Re} G_{k, \omega_n}(k) (v_+) \cos kx_- \quad (3.8)$$

These equations are expressed in relative coordinates: $v_- = v - v'$, $x_- = x - x'$ and centrix coordinates: $v_+ = \frac{v+v'}{2}$ assuming spatial homogeneity. The propagator $G_{k, \omega}(v)$ to be used in 3.7 is the usual renormalized propagator given by:

$$G_{k, \omega}(v) = \int_0^\infty \exp(i(\omega - kv)\tau) \exp\left(-\frac{k^2(v)D}{3}\tau^3\right) d\tau \quad (3.9)$$

Dupree¹⁷ was the first to show that equation (3.7) predicts that initially nearby trajectories will diverge exponentially in time. He considered the source free solution to eqn. (3.7), $G_2(x_-, v_-; v_+, t)$, and generated a closed set of equations for the moments $\langle x_-(t)^n v_-(t)^m \rangle \equiv \int dx_- \int dv_- x_-^n v_-^m G_2$. The solution of the resulting equations showed that $\langle x_-(t)^2 \rangle$, $\langle v_-(t)^2 \rangle$ increased from their initial values as $\exp\left(\left[4k_+^2 D^{Q.L.}\right]^{1/3} t\right)$. A more precise contact between the dynamical definition has been made by Rechester, Rosenbluth and White.⁶ Since the quantity h as defined by Eq. 2.14 is numerically observed to be independent of initial conditions, these authors argue that it is a statistical quantity. More explicitly they invoke ergodic properties which permit them to replace the time average

$$\frac{1}{t} \int_0^t \frac{\partial}{\partial t'} \ln \left[\frac{d(t')}{d0} \right] dt' = \frac{1}{t} \ln \frac{d(t)}{d0}$$

by the statistical average $\frac{\partial}{\partial t} < \ln \frac{d(t)}{d0} >$. By writing the quasilinear equation corresponding to the two-point, one-time correlation function they find:

$$h^{Q.L.} = \frac{\partial}{\partial t} < \ln \frac{d(t)}{d0} > = 0.36 \left(k_+^2 D^{Q.L.} \right)^{1/3} \quad (3.10)$$

These authors observed an excellent agreement between the statistical prediction $h^{Q.L.}$ and the experimental value h_{exp} obtained from numerical integration of the orbits (Fig. 5).

Thus there is excellent agreement between the predictions of the quasilinear statistical theories for velocity diffusion (i.e., $D_{\text{exp}} \approx D^{Q.L.}$) and the corresponding predictions for the Kolmogorov entropy (i.e., $h_{\text{exp}} \approx h^{Q.L.}$) in the case of prescribed quasigaussian random fields when the conditions (3.3) for the validity of quasilinear theory are satisfied. In Section 4 we will show there is no corresponding agreement between standard statistical theory and the results of certain self-consistent computer simulations.

We complete this section by pointing out the following question: why the results of an ensemble averaged statistical theory can accurately describe the results of a single numerical experiment; i.e., a single realization of the ensemble. The answer is really not obvious as it is sometimes considered to be; the statistical quantities $D^{Q.L.}$ and $h^{Q.L.}$ clearly do not depend on the phases, but the results D_{exp} and h_{exp} of one-numerical experiment with a given assignment of the phases could in principle depend on the phases. The simulations described above are essentially spatially homogeneous and explicitly calculated only spatial averages. In particular, the spatially averaged phase space distribution $\bar{f}(v,t)$ is defined as

$$\bar{f} = \frac{1}{L} \int_0^L dx f(x,v,t) \quad .$$

Pesme and Brisset¹⁸ have considered the following situation: for each realization the mode amplitudes $|E_k|$ form a smooth function of k and the phases θ_k are randomly distributed and independent of each other. For such a given realization they have shown that \bar{f} obeys the quasilinear equation (3.5) with $\langle f \rangle$ replaced by \bar{f} if the conditions (3.3) are satisfied. More precisely they have shown that the corrections to this equation are statistically small under these conditions. The condition $k_c \gg 1$ is easily shown to be equivalent to the following condition on the length of the system:

$$L/v \gg h^{-1} \quad \text{or} \quad L \gg \ell_h \equiv v \tau_h \quad (3.11)$$

In this case during the transit time of a particle of velocity v through the system it sees many independent partitions of the system which are equivalent to independent realizations. This is because on one hand, after a time h^{-1} the particle has lost all memory of its initial conditions. On the other hand, the correlation length of the electric field fluctuations is $\ell_c \approx 1/\Delta k$ and when the conditions 3.3 are satisfied, we have $\ell_c \ll \ell_h$ so that the electric field in each partition is independent. Therefore spatial averaging is then equivalent to ensemble averaging.

COMPARISON OF SELF-CONSISTENT NUMERICAL SIMULATIONS WITH STANDARD QUASILINEAR THEORY

Here we will review the simulations of the bump on tail instability in a one-dimensional, one-component plasma which were carried out and analyzed by Adam, Laval and Pesme.¹⁹ The conditions of these simulations were those commonly thought to be in the regime of validity of quasilinear theory.²⁰ In fact, the simulations did not follow the quasilinear predictions and this failure is now attributed to the breakdown of the random phase assumption in a self-consistent plasma.

The fact that self-consistency results in qualitatively and quantitatively different behavior than in the imposed field case has also been verified in a rather different physical situation by the simulations of Bezzerides, Gitomer and Forslund.²¹ These authors investigated the production of hot electrons in resonance absorption of light. Detailed consideration of this case is not possible here and we refer the reader to this work.

In the self-consistent problem the electric field is, of course, determined by the charge density of all the particles in the system. Poisson's equation is written as:

$$\nabla \cdot E(x,t) = 4\pi q \int dv f(x,v,t) - 4\pi q N \quad (4.1)$$

where f is a smooth function¹⁴ and N is the ion density. The Vlasov-Poisson system of equations thus becomes a nonlinear field theory. In the self-consistent case the field $E(x,t)$ cannot be specified, but evolves dynamically from the initial conditions ($f_0(x,v) = f(x,v,t=0)$).

The bump on tail instability develops from the initial conditions shown in Fig. 6a. Here is shown the initial spatial averaged distribution $f(v,t=0) = f_{k=0}(v,t=0)$. The phases of $f_k(v,t=0)$ are randomly chosen from one mode to the next. The fourier amplitudes $E_k(t=0)$ for $k \neq 0$ are computed from Poisson's equation, i.e., $E_k(t=0) = (4\pi q/ik) \int dv f_k(v,t=0)$ but are not shown. The initial amplitudes $|E_k(t=0)|$ are shown schematically in Fig. 6.a.

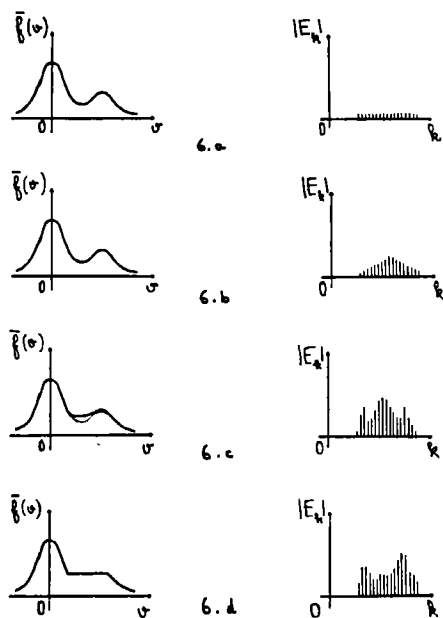


Fig. 6. Time evolution of the space-averaged distribution function $f(v,t)$ and of the electric field made amplitude $|E_k|(t)$, a. initial conditions, b. earlier stage: growth of $|E_k|$ and no modification of f , d. stationary final state.

Fig. 6

It is well-known that the linear dispersion relation derived from $\epsilon_0(k,\omega) = 0$ where $\epsilon_0(k\omega)$ is the linear dielectric function predicts the familiar Landau growth rate for Langmuir waves provided $\gamma_k^{Q.L.}/\omega_p \ll 1$:

$$\frac{\gamma_k^{Q.L.}}{\omega_k} = \frac{\pi \omega_p^2}{2Nk|k|} \left[1 - \frac{k}{\omega_k} \frac{d\omega_k}{dk} \right] \left(\frac{\partial \bar{f}}{\partial v} \right)_{v=\omega_k/k} \quad (4.2)$$

Particles with velocities v near the phase velocity $v_\phi(k) = \omega(k)/k$ of the wave see essentially a static field and interact strongly with the wave. If $(\partial_v \bar{f})_{v=v_\phi} > 0$ then the waves grow because there is energy transfer from

particles to the wave. For the initial condition on \bar{f} chosen by Adam et al., there is a band of wave numbers whose growth rates ($\gamma_k^{Q.L.} > 0$) are very close in value and for short times a broad spectrum of waves grows up as shown in Fig. 6b. At later times (Fig. 6c) the spectrum becomes noisy and uneven but still broad and begins to cause diffusion of particles in velocity which tends to flatten out \bar{f} in the unstable region of velocities. At longer times (Fig. 6d) \bar{f} becomes essentially flat in this region, the waves cease to grow and a broad but noisy quasistationary spectrum of waves results. Except for very short times the spectrum has grown to a level for which $k_s \gg 1$ and the condition for stochasticity expected for an imposed field is exceeded. Also, throughout the simulation the condition for quasilinear theory $\hbar^{-1} = \tau_h \ll \tau_{\bar{f}}$ is well satisfied.

The comparison between the numerical results and the quasilinear prediction for the growth rate for a typical unstable mode is shown in Fig. 7. The numerical results for the ensemble average over ten simulation runs (solid line) lie significantly above the predicted quasilinear growth rate (dashed line).

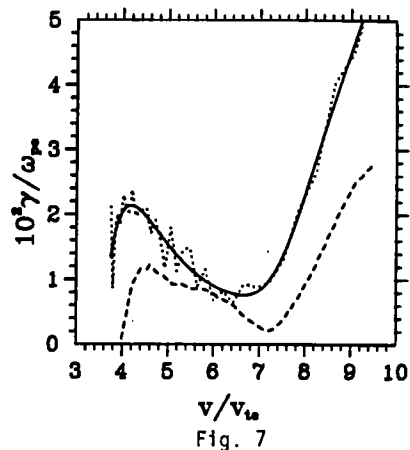


Fig. 7. Growth rates versus phase velocity: the dashed line is the quasilinear prediction, the dotted curve is obtained by numerical time derivation of the ensemble average of the mode energy and the solid curve is a least square fit of the preceding one (from Adam et al., Ref. 19).

Now recall the usual quasilinear equation for the time development of the mean intensity in mode k :

$$\partial_t \langle |E_k|^2 \rangle = 2 \gamma_k^{Q.L.} \langle |E_k|^2 \rangle + S_k \quad (4.3)$$

In the usual quasilinear theory the source term S_k which arises from mode-mode coupling effects is argued to be negligible when the conditions (3.3) are fulfilled. The coupled equations (4.3) with $S_k = 0$ and (3.5) constitute the usual quasilinear theory. Together they conserve particle number, momentum, and energy.

The simulations appear to show that the growth rate which should appear in (4.3) is greater than $\gamma_k^{Q.L.}$; in this case equation (3.5) must also be altered (e.g., $D \neq D^{Q.L.}$) in order to maintain conservation of energy. Therefore, we infer from the simulations that

$$\gamma_{\text{exp}}(k) \neq \gamma^{\text{Q.L.}}(k)$$

$$D_{\text{exp}} \neq D^{\text{Q.L.}}$$

Since of the conditions (3.3) for the validity of quasilinear theory both i) ($k_s \gg 1$) and ii) ($\hbar^{-1} \ll \tau_{<f>}$) are satisfied during the simulation we must conclude that iii) the quasis gaussian fluctuation assumption is not valid in the self-consistent field case.

This statement is not surprising in itself since we know that the self-consistent Poisson equation is nonlinear. As a result nongaussian electric field fluctuations will evolve from initially gaussian electric field fluctuations.

SELF-CONSISTENT STATISTICAL THEORIES OF PLASMA TURBULENCE

Introduction

The understanding of the self-consistent statistical description of Vlasov turbulence is in a new stage of development today and the study of this problem will certainly be the subject of future research in the nonlinear physics of plasmas.

In the prescribed electric field problem we were concerned with the motion of particles in random forces. The property for the electric field of being prescribed did not appear explicitly in the derivation of the equations of evolution for $\langle f \rangle$ and $\langle \delta f \delta f' \rangle$. These equations (eq. 3.6 and 3.7) actually reduce to two Fokker-Planck equations and they can be easily derived by computing the friction and diffusion coefficients corresponding to the motion of a particle in the random field. This is the reason why it has been thought for a long time that the statistical description of the particle motion in self-consistent Vlasov turbulence was the same as in the prescribed electric field case. We have, however, to keep in mind that a basic validity condition for quasilinear theory - as well as for the Fokker-Planck equation - is the quasis gaussian property of the electric field. In the stochastic acceleration problem this property was always imposed, e.g., through random phases, but in the Vlasov-Poisson case it is necessary to check the validity of this hypothesis.

We consider explicitly a self-consistent Vlasov turbulence situation for which the conditions 3.3i and 3.3ii are satisfied. We assume also the usual weak turbulence hypothesis that assigns N_m well-defined frequencies $\omega_n(k)$ to a wave number k . These frequencies have to be computed self-consistently through a nonlinear dispersion relation. For simplicity we will take $N_m = 1$. All the quantities such as k_s , ω_b , τ_c are now computed through the instantaneous values of the self-consistent electric field. Inequality 3.3i) is easily satisfied in unstable situations. Inequality 3.3ii) is usually considered as the validity condition of quasilinear theory and is fulfilled, for instance, in the warm beam-plasma instability.

We define as quasis gaussian theories the statistical descriptions of Vlasov turbulence which assume a priori that the self-consistent electric field has quasis gaussian properties. In this section we will first present the usual argument given for justifying this assumption. We will then show that the quasis gaussian theories are self-contradictory in the regimes where the nonlinearity plays a role. The reason for discrepancy will be exhibited through a perturbative calculation: the statistical properties of the self-consistent electric field

are not characterized by a unique correlation length $\ell_c = 1/\Delta k$ but also by the stochastic decorrelation length $\ell_h = v\tau_h$.

We close this Section by describing some of the features of the direct interaction approximation (DIA) which is a renormalized theory which takes into account certain nongaussian effects.

The quasis gaussian hypothesis

In the earlier derivations of quasilinear theory the random-phase assumption was usually implicitly made.^{22,23} The underlying reason has been made explicit by O'Neil²⁴ as follows:

Let us assume on physical grounds that well above the stochastic threshold the electric field is itself turbulent and is characterized by an autocorrelation length ℓ_c^{eff} . In a finite box of length L , the Fourier components of the electric field are defined as:

$$E_k(t) = \frac{1}{2\pi} \int_0^L dx E(x,t) \exp-ikx \quad \text{with } k = \frac{n2\pi}{L}.$$

If we now divide the interval $[0,L]$ into segments of length ℓ_0 where $\ell_c^{\text{eff}} \ll \ell_0 \ll L$, the electric field in each segment is independent of the field in other segments. Therefore E_k can also be written as

$$E_k = \frac{1}{2\pi} \sum_{j=1}^N \int_{(j-1)\ell_0}^{j\ell_0} dx E(x,t) \exp-ikx \quad \text{with } N = L/\ell_0.$$

Thus E_k appears to be the sum of many independent variables and by the central limit theorem E_k behaves as a Gaussian variable in the limit L/ℓ_0 very large.

Let us emphasize that the only hypothesis in this demonstration is the existence of an autocorrelation length ℓ_c^{eff} and that seems quite reasonable in a turbulent situation.

However, O'Neil observed²⁴ that the argument outlined above cannot be used to justify the quasis gaussian assumption (which he calls the random-phase approximation). This assumption would say, for example, that the $\sum_p \langle E_{k-q} E_q E_{p-k} E_{-p} \rangle$ can be replaced by $2\langle |E_{k-q}|^2 \rangle \langle |E_q|^2 \rangle$. The neglected terms are²⁴ individually of order $(\ell_0/L)^3$ whereas the two terms retained by assumption are of order $(\ell_0/L)^2$. The sum over p , however, introduces a factor (L/ℓ_0) in the neglected terms; thus the quasis gaussian assumption is not justified by this argument. In Section (5.3) an explicit dynamical calculation due to Laval and Pesme of the corrections to the quasis gaussian assumption will be reviewed.

The quasis gaussian theories

On the basis of the previous section, the quasigaussian theories assume a priori that the last condition iii) of 3.3 is satisfied - i.e., the electric field behaves as a quasigaussian process.

The scheme of these theories is the following:

- 1) Two Fokker-Planck equations describe the evolution of $\langle f \rangle$ and $\langle \delta f \delta f' \rangle$, namely the quasilinear equation 3.6 and the Dupree equation 3.7. The friction and diffusion coefficients involve the spectral density $\langle |E_k|^2 \rangle$.
- 2) The spectrum $\langle |E_k|^2 \rangle$ is then computed by writing Poisson's equation as:

$$k^2 \langle |E_k|^2 \rangle = (4\pi)^2 q^2 \int dv dv' \langle \delta f \delta f' \rangle_k \quad (5.1)$$

We thus obtain a closed set of three equations for the three unknowns $\langle f \rangle$, $\langle \delta f \delta f' \rangle$, and $\langle |E_k|^2 \rangle$. However, it is necessary to go beyond this level in order to compute the irreducible correlation functions of the electric field of higher order than 2. This scheme would be consistent only if these correlation functions satisfy the definition of a quasigaussian process given in §2.

An indirect demonstration of the inconsistency of this scheme in the regime $(k^2 D)^{1/3} \gg \gamma_k^{Q.L.}$ has been given by Adam, Laval and Pesme.²⁵ They solved the Dupree equation 3.7 and by inserting its solution into Poisson's equation 5.1; they found a growth rate different from the quasilinear prediction. This result, combined with quasilinear equation 3.6 does not allow energy conservation. These authors conclude that the quasilinear theory is inconsistent in one dimension and in the regime $(k^2 D^{Q.L.})^{1/3} \gg \gamma_k^{Q.L.}$ and showed that the contribution of the 4-point irreducible correlation function is not negligible, in contradiction to the quasigaussian hypothesis.

Even though the quasigaussian scheme appears to be invalid, the solution of the Dupree equation 3.7 exhibits certain features worth being mentioned. In the

limit of inequalities 3.3i) and ii), and for $|k(v_+) v_-| \ll \sup (\gamma_k^{Q.L.}, (k^2 D^{Q.L.})^{1/3})$, and $|x_-/v_+| \ll \inf [(k(v_+)^2 D^{Q.L.})^{-1/3}, \gamma_k^{-1}]$, equation 3.7 becomes:

$$\begin{aligned} \left[\partial_t + v_- \partial_{x_-} - 2D^{Q.L.}(v_+) \partial_{v_-}^2 \right] \langle \delta f \delta f' \rangle &= \\ 2 D^{Q.L.}(v_+) \cos(k(v_+)x_-) \left[\partial_{v_+} \langle f \rangle \right]^2 & \\ - 2D^{Q.L.}(v_+) \cos(k(v_+)x_-) \partial_{v_-}^2 \langle \delta f \delta f' \rangle & \end{aligned} \quad (5.2)$$

The last term of the R.H.S. of Eq. 5.2 is called the mode-coupling term since it couples together the k and $k-k(v_+)$, $k + k(v_+)$ Fourier components of $\langle \delta f \delta f' \rangle$.

The ratio $R = (k^2 D^{Q.L.})^{1/3} / \gamma_k^{Q.L.}$ turns out to be a measure of the nonlinearity and can be considered as a Reynolds number. For $R \ll 1$ the mode-coupling term is negligible and the quasilinear results are recovered. For $R \gg 1$ the mode coupling terms can be shown to be of the same order of magnitude as the other terms. The mode coupling terms give a contribution for $v \approx v' \approx \omega_k/k$ and generate

harmonics coupled together through the resonant particles. As a result, the correlation function $\langle \delta f \delta f' \rangle$ is found to have the following convenient harmonic expansion

$$\langle \delta f(x, v, t) \delta f(x', v', t') \rangle = \sum g_n(v_-) \exp[i n k(v_+)(x-x')] \exp[-in\omega_{|k(v_+)|}(t-t')] \quad (5.3)$$

In the regime $R \gg 1$, where the nonlinearity plays a role, the correlation function is thus the sum of an infinite number of harmonics coupled together. When inserted into Poisson equations the contribution coming from this harmonic coupling appears as a source term radiating coherently. We can, therefore, say that the nonlinearity reinforces the coherence of the solution. As a consequence, Adam, Laval, and Pesme indeed have found an increase of the growth rate compared to the quasilinear result. This feature is generalized in a self-consistent way in the DIA.

The Dupree "clump theory" likewise predicts an enhanced radiation due to macroscopic granulation of the phase space density - a sort of enhanced discreteness noise due only to nonlinear processes in the continuous Vlasov approximation. The Dupree theory applies to a "strong" turbulence limit in which no definite normal mode frequency exists for a given k . Therefore $k(v) = \omega_{k(v)}/v$ is not defined in this limit.

The results of Adam, Laval and Pesme apply in the limit of quasilinear ordering (eq. 3.3i and 3.3ii) when $v_k/\omega_k \ll 1$ so that a definite frequency is associated with a definite k . Particles of velocity v interact strongly with waves in the spectrum of wave number $k(v)$ such that particles tend to be periodically correlated with the separation $2\pi/k(v)$. Physically, the particles are more likely to be found near the troughs of these resonant waves.

The quasigaussian hypothesis revisited

In this subsection we will show in a perturbative way how the mode-coupling terms generate a nongaussian electric field. We will follow essentially the arguments of Laval and Pesme.²⁶

Let $\bar{f}_0(v, t)$ the space-averaged distribution function and define δf as $\delta f = f(x, v, t) - \bar{f}_0(v)$. Following Kadomtsev, the equation for δf can be written in Fourier space as:

$$\delta f_{k,w} = g_{kw}^0 [E_k f_0]_w + g_{kw}^0 \sum_{k' \neq 0} \int d\omega' E_{k-k', w-\omega'} f_{k', \omega'}$$

$$\text{where } g_{kw}^0 = \frac{(q/m)}{i[\omega - kv + i\alpha]} \frac{\partial}{\partial v},$$

in which α provides a contour prescription for v -integration.

Equation 5.5 can be solved iteratively beginning with $\delta f_{k,w}^{(1)} = g_{kw}^0 [E_k f_0]_w$ in a functional series in powers of E . This series is then inserted into Poisson's equation $ikE_{kw} = 4\pi q \int \delta f_{kw}(v) dv$ and yields a nonlinear equation for E_{kw} of the following form:

$$\varepsilon(k, \omega) E_{kw} = \mu_2(k, \omega) E E + \mu_3(k, \omega) E E E + \dots \quad (5.6)$$

where ε is the quasilinear dispersion relation:

$$\varepsilon(k, \omega) = 1 + i \frac{4\pi q}{k} \int dv g_{kw}^0 \bar{f}_0 \quad (5.7)$$

and

$$\mu_2(k, \omega) EE = \sum_{k'} \int d\omega' V_{kw, k'w'} E_{k'w'} E_{k-k' \omega-\omega'} \quad (5.8a)$$

$$\mu_3(k, \omega) EEE = \sum_{k'} \sum_{k''} \iint d\omega' d\omega'' V_{kw, k'w', k''w''} E_{k'w'} E_{k''w''} E_{k-k'-k'' \omega-\omega'-\omega''} \quad (5.8b)$$

with

$$V_{kw, k'w'} = \frac{q}{ik\varepsilon_0} \int g_{kw}^0 g_{k-k'}^0 \bar{f}_0 dv \quad (5.9)$$

$$V_{kw, k'w', k''w''} = \frac{q}{ik\varepsilon_0} \int g_{kw}^0 g_{k-k'}^0 g_{k-k'-k''}^0 \bar{f}_0 dv$$

Equation 5.7 can be solved iteratively by setting

$$E = E^{(1)} + E^{(2)} + E^{(3)} \dots \quad (5.10)$$

The lowest order gives $\varepsilon E^{(1)} = 0$. This is the approximation of quasilinear theory in which the imaginary part of the root of $\varepsilon(k, \omega)$ determines the quasilinear growth rate $\gamma_k^{Q.L.}$; for simplicity in the following we set $\gamma_k^{Q.L.} = \gamma_k$. We have $E_k^{(1)}(t) = E_k(0) \exp[-i\omega_k t + \gamma_k t]$. We assume $E_k(0)$ as a random phased field. The mode-coupling terms will generate nongaussian correlations:

$$\varepsilon E^{(2)} = \mu_2 E^{(1)} E^{(1)} \quad (5.11a)$$

$$\varepsilon E^{(3)} = \mu_2 E^{(2)} E^{(1)} + \mu_2 E^{(1)} E^{(2)} + \mu_3 E^{(1)} E^{(1)} E^{(1)} \quad (5.11b)$$

We now consider the 4-point correlation function $\langle E_{k_1+k_2+k_3} E_{-k_1} E_{-k_2} E_{-k_3} \rangle$.

On substitution of the expansion (5.10) for each E_k we find first of all the usual terms $\langle E_{\Sigma k_i}^{(1)} E_{-k_1}^{(1)} E_{-k_2}^{(1)} E_{-k_3}^{(1)} \rangle$ which, because of the gaussian assumption

for $E^{(1)}$ can be cast into a sum of $\langle E^{(1)} E^{(1)} \rangle \langle E^{(1)} E^{(1)} \rangle$. Let us now compute

the nongaussian contributions. At the lowest order we obtain the sum of four terms like $\langle E_{k_j}^{(3)} E_{k_j}^{(1)} E_k^{(1)} E_{k_{k_l}}^{(1)} \rangle$ with $\Sigma k_m = 0$. The terms $E^{(2)}$ are indeed smaller by a factor γ_k/ω_{pe} since $\varepsilon(\omega, k)$ in Eq. 5.11a has to be taken at frequency $\omega = 0$ or $\pm 2\omega_{pe}$ and cannot be made small.

From Eq. 5.11b we find

$$E_k^{(3)}(t) = \sum_{k'} \sum_{k''} \frac{V_{k\Sigma\Omega, k'\Omega_{k'}, k''\Omega_{k''}}}{\varepsilon(k, \Sigma\Omega)} E_{k'}^{(1)}(t) E_{k''}^{(1)}(t) E_{k-k'-k''}^{(1)}(t) \quad (5.12)$$

where $\Omega_k = \omega_k + i\gamma_k$ and $\Sigma\Omega = \Omega_{k'} + \Omega_{k''} + \Omega_{k-k'-k''}$. After substitution of this expression into $\langle E^{(3)} E^{(1)} E^{(1)} E^{(1)} \rangle$ and use of the gaussian assumption for $E^{(1)}$, we obtain the following expression:

$$\begin{aligned} & \langle E^{(3)}(t_1) E^{(1)}(t_2) E^{(1)}(t_3) E^{(1)}(t_4) \rangle \\ & \quad \substack{k_1+k_2+k_3 \\ -k_1 \quad -k_2 \quad -k_3} \\ & = \left(\frac{g}{m}\right)^2 \langle |E_{k_1}^{(1)}(0)|^2 \rangle \langle |E_{k_2}^{(1)}(0)|^2 \rangle \langle |E_{k_3}^{(1)}(0)|^2 \rangle \\ & \quad \exp -i\omega_{k_1}(t_1-t_2) \exp -i\omega_{k_2}(t_1-t_3) \exp -i\omega_{k_3}(t_1-t_4) \\ & \quad \exp \gamma_{k_1}(t_1+t_2) \exp \gamma_{k_2}(t_1+t_3) \exp \gamma_{k_3}(t_1+t_4) \\ & \cdot \left\{ \frac{2k_1^2}{\left[(2k_1+k_2+k_3) \frac{\omega_{k_1}}{k_1} - 2i\nu \right]^3 \left[(k_1+k_2) \frac{\omega_{k_1}}{k_1} - i\nu \right]} \right\} \\ & + \text{permutations } (k_1, k_2, k_3) \quad , \text{ with } \nu = \gamma_{k_1} \end{aligned} \quad (5.13)$$

Here the matrix element V has been explicitly evaluated using 5.8b. The dielectric function has been taken to be on resonance, i.e., $\omega_{k_1} + \omega_{k_2} + \omega_{k_3} \simeq \omega_{k_1+k_2+k_3}$ which implies that frequencies and wave numbers have not all the same sign. We have also assumed that the dispersion of the roots ω_k is weak so

that $\varepsilon(\Sigma\Omega) \simeq \left(\frac{\partial \varepsilon}{\partial \omega}\right) 2i\gamma_k^{Q.L.}$. The expression 5.13 enables us to compute the contribution of the irreducible 4-point correlation function in the evaluation of the mode amplitudes. At the lowest order we have

$$\begin{aligned} & i \frac{\partial \varepsilon}{\partial \omega_k} \left[\partial_t - 2\gamma_k^{Q.L.} \right] \langle |E_k(t)|^2 \rangle \\ & = \langle E_{-k}^{(3)} \mu_3(\Omega_k, k) E^{(1)} E^{(1)} E^{(1)} \rangle + \langle E_k^{(3)} \mu_3(\Omega_{-k}, -k) E^{(1)} E^{(1)} E^{(1)} \rangle \end{aligned} \quad (5.14)$$

which leads to the result:

$$\langle |E_k| ^2(t) \rangle = \langle |E_k^{(1)}(t)|^2 \rangle \left[1 + \left[0.24(k^2 D)^{1/3} / \gamma_k \right]^6 \right] \quad (5.15)$$

The modification of the spectrum as given by the expression inside the bracket of Eq. 5.15 shows again that the mode coupling terms lead to an increase of the energy transfer from resonant particles towards waves. This result is strictly valid only for $\gamma_k^{Q.L.} \gg (k^2 D)^{1/3}$ but it clearly shows that the irreducible 4-point correlation function gives a non-negligible contribution as soon as $(k^2 D)^{1/3} \gtrsim \gamma_k^{Q.L.}$ which invalidates the quasigaussian assumption.

By Fourier transform of Eq. 5.13 we obtain the expression in real space of the 4-point correlation function

$$\begin{aligned} & \langle E^{(3)}(x_1, t_1) E^{(1)}(x_2, t_2) E^{(1)}(x_3, t_3) E^{(1)}(x_4, t_4) \rangle = \\ & \sum_{k_1, k_2, k_3} \langle E^{(3)}(t_1)_{k_1+k_2+k_3} E^{(1)}(t_2)_{-k_1} E^{(1)}(t_3)_{-k_2} E^{(1)}(t_4)_{-k_3} \rangle \\ & \exp i k_1(x_1-x_2) \exp i k_2(x_1-x_3) \exp i k_3(x_1-x_4) \end{aligned} \quad (5.16)$$

If the expression in the curly brackets of the R.H.S. of Eq. 5.13 was a smooth function of k_1 , k_2 and k_3 , then this correlation function would be essentially a product of three two-point correlation functions of the form

$$\phi(\xi, \tau) = \Sigma \langle |E_k(0)|^2 \rangle \exp -i[\omega_k \tau - k\xi]$$

For correlations along the unperturbed orbits $\xi = v\tau$, this has the usual correlation time $\tau_c = |\Delta k(v-v_g)|^{-1}$ discussed previously.

On the contrary, however, the expression in curly brackets of the R.H.S. of Eq. 5.13 is a peaked function for $v|k_1+k_2| \lesssim v$. In the regime $R = (k^2 D)^{1/3} / \gamma_k \gg 1$, Eq. 5.15 shows that the series expansion in power of $E^{(1)}$ is not convergent. The free propagator g may then be replaced by the renormalized propagator g^r

$$g_{\omega, k}^r = (-q/m) \int_0^\infty d\tau \exp -i(\omega - kv)\tau \exp -\left[\frac{(k^2 D^{Q.L.}) \tau}{3} \right]^3 \frac{\partial}{\partial v} \quad (5.17)$$

In the estimation of order of magnitudes it is sufficient to replace $v = \gamma_k^{Q.L.}$ by $v = (k^2 D^{Q.L.})^{1/3}$. The nongaussian correlation function is, therefore, characterized by the time $(k^2 D^{Q.L.})^{-1/3}$ which is of the order of the inverse of Kolmogorov entropy $h^{Q.L.} = 0.36(k^2 D^{Q.L.})^{1/3}$ encountered earlier. Since $(k^2 D)^{-1/3} \gg \tau_c$ we see that self-consistency introduces a tendency toward self-organization of the turbulent system.

A detailed diagrammatic expansion leads to the conclusion that in the regime $(k^2 D)^{1/3} \gg \gamma_k$ all the irreducible correlation functions contribute at the same order of magnitude as the quasilinear term. This result invalidates the theories which consist of a simple renormalization of the propagator such as developed by

Rudakov and Tsytovich, Choi and Horton.²⁷ In the regime defined by 3.3i) and ii) they lead to the quasilinear result since they don't contain the mode coupling term of Eq. 5.2. It can be shown²⁸ that a partial summation of the neglected irreducible correlation functions leads to the Dupree equation 3.7. Therefore, we may summarize the different levels of renormalization:

1. WEAK TURBULENCE THEORY - This makes the usual quasilinear prediction but suffers from the defect of secular divergences arising in the neglected irreducible correlation functions.

These secularities are removed by renormalization of the propagator which leads to:

2. SIMPLY RENORMALIZED THEORIES - These again make the usual quasilinear prediction but still suffer from non-negligible contributions from neglected irreducible correlation functions.

A partial summation of the irreducible correlation functions leads to:

3. DUPREE'S THEORY - This theory predicts a zero order modification of the quasilinear result coming from mode coupling terms. However, this theory considers only part of the irreducible correlation functions and is not properly self-consistent.

SELF-CONSISTENT STATISTICAL DESCRIPTIONS AND THE DIA

It appears from the considerations above that most familiar statistical theories of Vlasov turbulence until recently, have been quasigaussian theories in the sense that they explicitly or implicitly make the assumption of quasigaussian electric field correlations. A self-consistent statistical description of Vlasov turbulence must take into account the relationship between perturbations in the averaged phase space distribution $\langle f_1 \rangle \equiv \langle f(v_1, x_1, t_1) \rangle$ and perturbations in the mean electric field $\langle E_1 \rangle = \langle E(x_1, t_1) \rangle$ as well as the relationship between the fluctuations in these quantities: $\delta f_1 = f_1 - \langle f_1 \rangle$ and $\delta E_1 = E_1 - \langle E_1 \rangle$. These relationships are, of course, imposed by Poisson's eqn. (4.1). For the discussion here we follow closely the results of reference 29.

From the Vlasov eqn. (32) the well-known equations for the mean distribution function is easily found:

$$\left[\partial_{t_1} + v_1 \partial_{x_1} \right] + \frac{q}{m} \langle E_1 \rangle \partial_{v_1} \langle f_1 \rangle = -\frac{q}{m} \partial_{v_1} \langle \delta E_1 \delta f_1 \rangle + \eta_1 \quad (5.18)$$

This couples $\langle f \rangle$ to the fluctuation correlation function $\langle \delta E \delta f \rangle$. The averaged Poisson equation is

$$\nabla_1 \cdot \langle E_1 \rangle = 4\pi q \int dv_1 \langle f_1 \rangle - 4\pi q N \quad (5.19)$$

We have added an artificial phase space source density $\eta_1 = \eta(x_1, v_1, t_1)$ which we can consider as a probe of the system. Suppose for $\eta = 0$ the system of equations has a well-defined solution $\langle f_1 \rangle, \langle E_1 \rangle$; we then add a local phase space source perturbation $\delta \eta$ which changes $\langle f_1 \rangle$ to $\langle f_1 \rangle + \delta \langle f_1 \rangle$ and $\langle E_1 \rangle$ to $\langle E_1 \rangle + \delta \langle E_1 \rangle$.

The increments δ are functional differentials such as those used in the calculus of variations of continuous fields. A relationship between the increments $\delta \langle f \rangle, \delta \langle E \rangle$ and $\delta \langle \eta \rangle$ is readily found from eq. (5.18) to be of the form

$$G_{12}^{-1} \delta \langle f_2 \rangle + \frac{q}{m} \partial_{v_1} \Gamma_{12} \delta \langle E_2 \rangle = \delta \eta_1 \quad (5.20)$$

By taking the functional differential of eqn. (5.18) it is easy to identify the coefficients as

$$G_{12}^{-1} = (\partial_{t_1} + v_1 \partial_{x_1} + \frac{q}{m} \langle E_1 \rangle \partial_{v_1}) \delta(1-2) + \frac{q}{m} \partial_{v_1} \frac{\delta \langle \delta E_1 \delta f_1 \rangle}{\delta \langle f_2 \rangle} \quad (5.21)$$

$$\Gamma_{12} = \langle f_1 \rangle \delta(1-2) + \frac{\delta \langle \delta E_1 \delta f_1 \rangle}{\delta \langle E_2 \rangle} \quad (5.22)$$

To simplify writing the equations we are using a summation convention where, for example, $G_{12}^{-1} \delta \langle f_2 \rangle \equiv \int dv_2 \int dx_2 \int dt_2 G_{12}^{-1} \delta \langle f_2 \rangle$, etc. and $\delta(1-2) \equiv \delta(v_1 - v_2) \delta(x_1 - x_2) \delta(t_1 - t_2)$. The last terms on the right hand side of eqns. (5.21) and (5.22) contain functional derivatives of the correlation function $\langle \delta E_1 \delta f_1 \rangle$ with respect to changes in the mean values $\langle f \rangle$ and $\langle E \rangle$, respectively. These functional derivatives completely describe the renormalization of the response functions G_{12} and Γ_{12} .

The response function G_{12} is called the "quasiparticle propagator"; without renormalization it is clear from eqn. (5.21) that G_{12} simply describes particle propagation along unperturbed orbits (in the mean field $\langle E_1 \rangle$ which is zero in a homogeneous system.) The renormalization term takes into account the self-consistent response of the fluctuating turbulent "background" through which the particle propagates. In an imposed fluctuation limit this propagator reduces to the familiar Dupree renormalized propagator given in eqn. 3.9.

The response function $\partial_{v_1} \Gamma_{12}$ is called the "quasiparticle potential source."

It takes into account the effective source density $\delta \eta$ which must arise when there is a change in the mean electrostatic field $\delta \langle E \rangle$ in order to keep $\langle f \rangle$ constant. Without renormalization, this term is simply proportional to $\partial_{v_1} \langle f_1 \rangle$.

The renormalization of this quantity is not familiar but was actually considered by Kadomtsev³⁰ in Section III of his 1965 book (equation 21a). (Kadomtsev's renormalized theory has some elements in common with the DIA, but differs in several important respects related to self-consistency.)

We next consider the fluctuations. By subtracting eq. (5.18) from eqn. (3.2) and noting the definitions (5.21) and (5.22) we can derive the exact equation for δf_1

$$\delta f_1 = \overbrace{(q/m) G_{12} \partial_{v_2} \Gamma_{23} \delta E_3}^{\delta f_1^{\text{coh}}} + \overbrace{(q/m) G_{12} \partial_{v_2} \left[\delta f_2 \delta E_2 - \langle \delta f_2 \delta E_2 \rangle - \frac{\delta \langle \delta f_2 \delta E_2 \rangle}{\delta \langle f_3 \rangle} \delta f_3 - \frac{\delta \langle \delta f_2 \delta E_2 \rangle}{\delta \langle E_3 \rangle} \delta E_3 \right]}^{\delta f_1^{\text{inc}}} \quad (5.23)$$

We have indicated that δf_1 can be separated into a coherent part proportional to the electrostatic fluctuation δE and an incoherent part. In a spatially homogeneous plasma it is easy to see that the Fourier component δf_k^{coh} is proportional to δE_k . For the incoherent part terms like $\delta f_1 \delta E_1$ involve Fourier convolutions such as $\sum_{k'} \delta f_{k-k'} \delta E_{k'}$ (with $\delta E_{k=0}=0$) and thus involves the coupling of different Fourier modes.

If the expression (5.23) is inserted into Poisson's equation $\nabla_1 \cdot \delta E_1 = 4\pi q \int dv_1 [\delta f_1^{\text{coh}} + \delta f_1^{\text{inc}}]$ one can rearrange the coherent contribution to write

$$\nabla_1 \cdot \epsilon_{12} \delta E_2 = 4\pi q \int dv_1 \delta f_1^{\text{inc}} \quad (5.24)$$

where ϵ_{12} is the dielectric operator²⁹

$$\epsilon_{12} = \delta(t_1 - t_2) \delta(x_1 - x_2) + \frac{4\pi q^2}{m} \nabla_1 \cdot \int dv_3 \int dx_3 \frac{\delta(t_1 - t_3)}{|x_1 - x_3|} G_{34} \partial_{v_4} \Gamma_{42} \quad (5.25)$$

The coherent part has produced the susceptibility function in the dielectric operator. For a homogeneous stationary system, the Fourier coefficient of ϵ can be written

$$\epsilon(k, \omega) = 1 + (4\pi q^2 / mk^2) \int dv_3 \int dv_4 G_{v_3 v_4}(k, \omega) ik \cdot \partial_{v_4} \Gamma_{v_4}(k, \omega)$$

We see from eqn. (5.24) that the charge density associated with δf^{inc} acts as an incoherent source for the electrostatic fluctuation. For example, the two-point electrostatic correlation function can be written as

$$\langle \delta E_1 \delta E_{1'} \rangle = \partial_{x_1} (\nabla_1^2 \epsilon_{12})^{-1} (4\pi q) \int dv_2 \int dv_2' \langle \delta f_2^{\text{inc}} \delta f_{2'}^{\text{inc}} \rangle \cdot \partial_{x_1'} (\nabla_{1'}^2 \epsilon_{1'2'}) \quad (5.26a)$$

in terms of the inverse dielectric operator. Or

$$\langle |\delta E|^2 \rangle_{kw} = \frac{(4\pi q)^2}{k^2 |\epsilon(k\omega)|^2} \int dv \int dv' \langle \delta f_v^{\text{inc}} \delta f_{v'}^{\text{inc}} \rangle_{k, \omega} \quad (5.26b)$$

for the spectral intensity in a homogeneous, stationary case.

The two-point correlation function for the phase space distribution can likewise be written as

$$\begin{aligned} \langle \delta f_1 \delta f_{1'} \rangle &= (q^2 / m^2) G_{12} \partial_{v_2} \Gamma_{23} G_{1'2'} \partial_{v_2'} \Gamma_{2'3'} \langle \delta E_3 \delta E_{3'} \rangle \\ &+ \left[\delta(1-2) \delta(1'-2') - \delta(1-2) P_{1'2'} - P_{12} \delta(1'-2') \right] \langle \delta f_2^{\text{inc}} \delta f_{2'}^{\text{inc}} \rangle \end{aligned} \quad (5.27)$$

where P_{12} is a "polarization" operator

$$P_{12} = (4\pi q^2 / m) G_{13} \partial_{v_3} \Gamma_{34} (\nabla_4^2 \epsilon_{42})^{-1} \quad (5.28)$$

The first term on the right side is the correlation $\langle \delta f_1^{\text{coh}} \delta f_{1'}^{\text{coh}} \rangle$ while the first term in square brackets arises from $\langle \delta f_1^{\text{inc}} \delta f_{1'}^{\text{inc}} \rangle$. The second and third term in

square brackets arise respectively from $\langle \delta f_1^{\text{inc}} \delta f_1^{\text{coh}} \rangle$ and $\langle \delta f_1^{\text{coh}} \delta f_1^{\text{inc}} \rangle$. In writing these terms in terms of the operator P we have used Poisson's equation in the form of eq. (5.24) to eliminate δE in terms of δf^{inc} . By doing this we have expressed correlation functions of the form $\langle \delta E \delta E \delta f \rangle$ in terms of $\langle \delta f_1^{\text{inc}} \delta f_1^{\text{inc}} \rangle$ which can be seen from the definition of δf^{inc} , eq. 5.23, to involve 4-point and higher correlation functions. The vanishing of the cross correlation terms is made explicitly in the theory of Dupree (ref. 17 eqn. (21)).

The equations written in this section up to this point are formally exact. We see that the system is not closed since the correlation function $\langle \delta f_1^{\text{inc}} \delta f_2^{\text{inc}} \rangle$ involves three-point and four-point correlation functions.

An approximation which closes these equations called the "Direct Interaction Approximation," DIA, has attracted renewed interest. This type of theory, first applied by Kraichnan to Navier-Stokes turbulence,³¹ treats certain non-gaussian correlations in a nonperturbative way. This approach was reasonably successful for moderate Reynold's numbers but failed to accurately describe the inertial range for high Reynold's numbers. The failure was due to the inaccurate treatment of the convection of short scale turbulence by the long scale motion. The DIA was written down for the Vlasov-Poisson system by Orszag and Kraichnan³² in 1967. This work received very little attention until recently when DuBois and Espedal³³ showed how to "factor" the theory into single (quasi) particle and collective responses. This exposed the existence in the DIA of new self-consistent fluctuation (or nonlinear polarization) contributions which were not found in the older quasigaussian imposed field theories. The new terms are proportional to the dielectric response ϵ^{-1} such as the terms in the polarization operator in eqn. (5.27); various estimates have shown these terms to be of the same order as the imposed field terms. The DIA based theory was also investigated by Krommes and coworkers³⁴ who also considered application to the guiding center model and to drift wave turbulence.³⁵

Here we will list the approximations to the set of exact equations listed above which are equivalent to the DIA:

1. The incoherent correlation function $\langle \delta f_1^{\text{inc}} \delta f_1^{\text{inc}} \rangle$ is computed to terms of quadratic order in the correlations using a quasigaussian approximation for the mixed four-point correlation function

$$\begin{aligned} \langle \delta f_1^{\text{inc}} \delta f_1^{\text{inc}} \rangle &\cong (q^2/m^2) G_{12} \partial_{v_2} G_{1'2'} \partial_{v_2'} [\langle \delta f_2 \delta E_2 - \langle \delta f_2 \delta E_2 \rangle \rangle] [\langle \delta f_2' \delta E_2' - \langle \delta f_2' \delta E_2' \rangle \rangle] \\ &\cong q^2/m^2 G_{12} \partial_{v_2} G_{1'2'} \partial_{v_2'} [\langle \delta E_2 \delta E_2' \rangle \langle \delta f_2 \delta f_2' \rangle + \langle \delta E_2 \delta f_2' \rangle \langle \delta f_2 \delta E_2' \rangle] \end{aligned} \quad (5.29)$$

Note that this is not at all just the quasigaussian approximation for the correlation function of four electric fields. The fluctuation δf contains, as a consequence of the incoherent contribution, δf^{inc} , in eqn. (5.23), terms of all orders in the electric field fluctuation δE . (These terms can be obtained from eqn. (5.23) by iteration starting with $\delta f_1^{(0)} = \delta f_1^{\text{coh}}$.) Thus the DIA contains

electric field correlations of all orders; in particular it produces in third order in (δE^2) the nongaussian correlations calculated by Laval and Pesme.

2. The inverse propagator G^{-1} and the response function Γ are computed to lowest (linear) order in the correlation functions.

To carry out the last step the functional derivatives of $\langle \delta E \delta f \rangle$ in eqs. (5.21) and (5.22) must be evaluated to this order. This can be accomplished by using an exact relation connecting the fluctuations which is easily derived from eq. (5.23):

$$\langle \delta f_1 \delta E_{1'} \rangle = \frac{q}{m} G_{12} \partial_{v_2} \Gamma_{23} \langle \delta E_3 \delta E_{1'} \rangle + 4\pi q \partial_{x_1'} (\nabla_{1'}^2 \epsilon_{1'2'})^{-1} \langle \delta f_1^{\text{inc}} \delta f_2^{\text{inc}} \rangle \quad (5.31)$$

This equation can also be used for the right-hand side of eq. (5.18) for the mean distribution. In computing the required functional derivatives of $\langle \delta f \delta E \rangle$ for self-consistent fluctuations it is necessary to consider the functional dependence on $\langle f \rangle$ or $\langle E \rangle$ of the dielectric response, ϵ^{-1} , which enters explicitly in the last term of eq. (5.31) and also in $\langle \delta E \delta E \rangle$ as shown explicitly in eq. (5.26). We can give here only the result of this calculation for G_{12}^{-1} for a homogeneous, stationary system:

$$G_{VV'}^{-1}(k, \omega) = -i(\omega - k \cdot v) \delta(v - v') + \frac{q^2}{m^2} \partial_{v'} \cdot \int dk' \int d\omega' \left\{ \langle \delta E \delta E \rangle_{k', \omega'} \right. \\ \left. \left[G_{VV'}(k - k', \omega - \omega') + \int d\bar{v} P_V(k - k', \omega - \omega') G_{VV'}^{-1}(k - k', \omega - \omega') \right] \right. \\ \left. + 4\pi q \frac{\langle \delta f_V \delta E \rangle_{k', \omega'}}{(k - k')^2 \epsilon(k - k', \omega - \omega')} \int d\bar{v} G_{VV'}^{-1}(k - k', \omega - \omega') \right\} \cdot \partial_{v'} \quad (5.32)$$

The last two terms in curly brackets are self-consistent fluctuation terms which arise from the functional dependence discussed above and do not occur in the quasigaussian field theories. These self-consistent fluctuation contributions have been estimated in various cases to be as important as the remaining term which appears in the quasigaussian field theories and produces the renormalized propagator in eq. 5.1. In the quasilinear ordering of $\tau_h \ll \tau_{cf}$ and $k_s \gg 1$ for a one-dimensional one-component plasma we have been able to show that the self-consistent fluctuation terms cannot be neglected. A similar equation results for the renormalized $\Gamma_V(k, \omega)$ which we will not discuss here. Further results are given in reference 29.

If the DIA approximation of eqn. (5.30) for $\langle \delta f^{\text{inc}} \delta f^{\text{inc}} \rangle$ is inserted into eq. (5.27) we obtain the following equation for the two-point correlation function

$$\langle \delta f_1 \delta f_{1'} \rangle = (q^2/m^2) \left[G_{12} G_{1'2'} - \overbrace{P_{13} G_{32} G_{1'2'} - G_{12} P_{1'3'} G_{3'2'}}^{\text{s.c.}} \right]$$

$$\begin{aligned}
 & \cdot \partial_{v_2} \partial_{v'_1} \left[\langle \delta E_2 \delta E_{2'} \rangle \langle \delta f_2 \delta f_{2'} \rangle + \overbrace{\langle \delta E_2 \delta f_{2'} \rangle \langle \delta f_2 \delta E_{2'} \rangle}^{\text{s.c.}} \right] \\
 & = (q^2/m^2) G_{12} \partial_{v_2} \Gamma_{23} G_{1'2'} \partial_{v'_1} \Gamma_{2'3'} \langle \delta E_3 \delta E_{3'} \rangle
 \end{aligned} \quad (5.34)$$

When the terms labelled s.c. are dropped in this expression and only the imposed field renormalization terms in G are retained, this equation can be shown to reduce to the Dupree equation for $\langle \delta f(x, v, t) \delta f(x' v' t) \rangle$ given in equation (3.7). The latter approximation is valid only for quasigaussian imposed electric field fluctuations; we can see explicitly in this case that the remaining contribution from $\langle \delta f_1^{\text{inc}} \delta f_1^{\text{inc}} \rangle$ in eqn. (5.34); i.e., $(q^2/m^2) G_{12} G_{1'2'} \delta_{v_2} \delta_{v'_1} \langle \delta f_2 \delta f_{2'} \rangle \langle \delta E_2 \delta E_{2'} \rangle$, reduces to the mutual diffusion (or mode coupling) term of equations (3.7) or (5.2). It is this term which produces the finite Kolmogorov entropy, h , discussed in Section 3. Thus there is a direct connection in this case between a finite level for the incoherent noise function $\langle \delta f_1^{\text{inc}} \delta f_1^{\text{inc}} \rangle$ (sometimes called the nonlinear noise) and the intrinsic stochastic behavior of the system.

The reduction of these DIA equations to a calculable form is very difficult even in the limit of quasilinear ordering. The harmonic expansion of eq. (5.3) still appears to be useful in the DIA. Using this technique we have been able to show that none of the self-consistent terms can be dropped even in quasilinear ordering when $R \gg 1$. Only the equations for the harmonic components $n = 0, \pm 1$ and 2 are significantly changed by the self-consistent terms. Work is still in progress in deducing the consequences of the DIA in the quasilinear limit.

The DIA is the natural first approximation in the renormalized perturbation theories of Kraichnan, Martin, Siggia and Rose (MSR)³⁶ and others. The general conditions for the convergence of the MSR expansion are completely unknown; probably only convergence in the sense of an asymptotic expansion is expected at best. It appears to be a possible, but extremely tedious, task to examine the size of the leading vertex corrections (in the sense of MSR) to the DIA. There is some hope that in quasilinear ordering these vertex corrections may be negligible. This hope is based on the following points: i) In the imposed field case the DIA theory becomes exactly quasilinear theory (QLT) in quasilinear ordering; the corrections in this case are of order $(\tau_h/\tau_{cf}) \ll 1$. ii) In the self-consistent field case the DIA differs from QLT in quasilinear ordering; the differences due to nonlinear polarization effects are the same order as the QLT terms. iii) If self-consistency also does not change the order of the vertex corrections, the corrections to the DIA will also be of order (τ_h/τ_{cf}) as in i).³⁷

To make a point of contact between this formalism and the results of Adam, Laval and Pesme, we write an equation for the time evolution of $\langle |\delta E|_k^2 \rangle$:

$$\partial_t \langle |\delta E|_k^2 \rangle = 2 \hat{y}_k \langle |\delta E|_k^2 \rangle + S_k \quad (5.35)$$

where

$$S_k = (4\pi q)^2 k^{-2} \int dv \int dv' \langle \delta f_v^{\text{inc}} \delta f_{v'}^{\text{inc}} \rangle_{k, \omega=\omega(k)} (\partial \epsilon / \partial \omega_k) \quad (5.36)$$

This can be shown to be an exact consequence of eq. (5.26a) provided $\hat{\gamma}_k < \hat{\omega}_k$ where $\hat{\omega}_k + i\hat{\gamma}_k$ is the zero of the dielectric function in the complex ω plane: $\epsilon(k, \hat{\omega}_k + i\hat{\gamma}_k) = 0$. In the limit of quasilinear ordering ($\tau_h < \tau_f$, $k_s \gg 1$) it is found that $\hat{\omega}_k$ and $\hat{\gamma}_k$ are given by their linear values: $\hat{\omega}_k^2 = \omega_p^2 + 3k^2 V_e^2$, $\hat{\gamma}_k = \gamma_k^{Q.L.}$. In conventional quasilinear theory it is assumed that S_k arises from mode-mode coupling effects of order $\langle |\delta E_k|^2 \rangle^2$ and is thus negligible. However, to account for the results of Adam, Laval and Pesme discussed in Section 4 it is necessary that to leading order

$$S_k = (\gamma_k^{\text{exp}} - \gamma_k^{Q.L.}) \langle |\delta E_k|^2 \rangle \quad (5.37)$$

where γ_k^{exp} is the observed growth rate. We see that S_k is related to the right hand side of eq. (5.14). Comparison can be made with the iterative calculation of Laval and Pesme (see Section 5.3) by iterating Eq. (5.34) starting with $\langle \delta f_1 \delta f_1 \rangle = [\text{right hand side of eq. (5.34)}]$ and inserting this formal expansion in powers of $\langle \delta E \delta E \rangle$ into eqs. (5.31) and (5.29). The combination of the last two equations then leads to a formal expansion of $\langle \delta f_1^{\text{inc}} \delta f_1^{\text{inc}} \rangle$ and also of S_k in powers of $\langle \delta E \delta E \rangle$. To terms formally of order $\langle \delta E \delta E \rangle^3$ this expansion agrees with the results of Laval and Pesme which was outlined in Section 5.3; because of wave-particle resonances for $\gamma_k < (k^2 D^{QL})^{1/3}$ the terms in S_k of this formal order appear to be actually of order $\langle |\delta E_k|^2 \rangle$, a result which is consistent with eq. (5.37). A complete solution of the DIA equations for $\langle \delta f_v^{\text{inc}} \delta f_v^{\text{inc}} \rangle_{kw}$ remains to be carried out. The self-consistent (nonlinear polarization) terms of the DIA do not appear to change the order in $\langle |\delta E_k|^2 \rangle$ of quantities such as S_k but they are numerically important and must be included, for example, to preserve energy conservation.

Conclusion

In the first part of this paper we have reviewed the results of intrinsic stochasticity theories in the case where the imposed electric field has random phase Fourier components. Whenever the mode amplitudes are well above the threshold for intrinsic stochasticity, an excellent agreement has been demonstrated between the numerical experiments and the quasilinear predictions for the diffusion coefficient and the Kolmogorov entropy.

In the second part we have considered the Vlasov-Poisson turbulence in which the electric field is self-consistently related to the particles motion through Poisson's equation. We have first observed a striking difference compared with the imposed electric field case: the numerical simulations exhibit a significant discrepancy with the quasilinear prediction. The origin of this discrepancy has then been shown to be a consequence of a strong enhancement of the wave-particle interaction. As a result, the statistical properties of the self-consistent electric field are characterized by a second correlation time $(k^2 D)^{-1/3}$ much longer than the usual correlation time τ_e . Self-consistency leads thus to a sort of self-organization of the turbulence. Dupree's theory now appears to be a partial summation of the irreducible correlation functions of the electric field which are neglected in the simply renormalized theories. It has, however, a remaining lack of consistency intrinsic to all quasis Gaussian

theories. We have then displayed the characteristic features of DIA applied to Vlasov turbulence and we have shown how this closure takes into account in a self-consistent way the strong wave-particle interaction and the resulting nongaussian properties of the electric field.

ACKNOWLEDGEMENTS

We are pleased to acknowledge helpful comments from Dr. J. C. Adam, Dr. G. Laval, Dr. H. Rose and Dr. T. O'Neill. Conversations with Dr. F. Doveil, Dr. D. Escande and Dr. A. Rechester are also acknowledged.

REFERENCES

1. Chirikov, B. V., Phys. Rep. 52, 263 (1979).
2. Escande, D. F. and Doveil, F., Phys. Lett. 83A, 307 (1981); Doveil, F. and Escande, D. F., Phys. Lett. 84A, 399 (1981); Escande, D. F. and Doveil, F., J. Stat. Phys. 26, 257 (1981).
3. Greene, J. M., J. Math. Phys. 9, 760 (1968). Lieberman, M. A. and Lichtenberg, A. J., Phys. Rev. A 5, 1852 (1972). Rosenbluth, M. N., Phys. Rev. Lett. 29, 408 (1972). Greene, J. M., J. Math. Phys. 20, 1183 (1979). Schmidt, G., Phys. Rev. A, 22, 2849 (1980). For references on applications see, e.g., Smith, G. R., and Kaufman, A. N., Phys. Fluids 21, 2230 (1978).
4. Rechester, A. B., and White, R. B., Phys. Rev. Lett. 44, 1586 (1980).
5. Rechester, A. B., Rosenbluth, M. N., and White, R. B., Phys. Rev. A 23, 2664 (1981). Cohen, R. H., and Rowlands, G., Lawrence Livermore National Laboratory preprint UCRL-85600, University of California (1981). Karney, C. F. F., Rechester, A. B., and White, R. B., Princeton Plasma Physics Laboratory Report P.P.P.L. 1752 (1981). Cary, J. R., Meiss, J. D., and Bhatta Charjee, A., Institute for Fusion Studies Report IFSR-6, Austin (1981).
6. Rechester, A. B., Rosenbluth, M. N., and White, R. B., Phys. Rev. Lett. 42, 1247 (1979).
7. Rechester, A. B., Rosenbluth, M. N., and White, R. B., in: Intrinsic Stochasticity in Plasmas, edited by G. Laval and D. Gresillon (Les Ed. de Physique, Orsay, 1979), p. 239.
8. Molvig, K., Freidberg, J. P., Potok, R., Hirshman, S. P., Whitson, J. C., and Tajima, T., Institute for Fusion Studies Report, Austin (1981).
9. Flynn, R. W., Phys. Fluids 14, 956 (1971).
10. Doveil, F., and Gresillon, D., in Proceedings of the International Congress on waves and instabilities in Plasma, Palaiseau, France 1977, p. 51.
11. Selected Papers on Noise and Stochastic Processes, edited by N. Wax (Dover, New York, 1954), p. 255.
12. Gary, S. P., and Montgomery, D., Phys. Fluids 11, 2733 (1968).
13. Graham, K. N., and Fejer, J. A., Phys. Fluids 19, 1054 (1976).

14. In the complete physical problem of a collection of discrete particles $f(x_0, v_0)$ (and therefore, $f(x, v, t)$) is a very singular object. We will consider here only the Vlasov approximation which assumes that $f(x_0, v_0)$ is a smooth function of x_0 and v_0 ; this is known to be valid if $n\lambda_D^3 \gg 1$ where n is the plasma density and λ_D the Debye length.
15. Sturrock, P. A., Phys. Rev. 141, 186 (1966).
16. Thomson, J. J., and Benford, G., J. Math. Phys. 14, 531 (1973).
17. Dupree, T. H., Phys. Fluids 15, 334 (1972).
18. Pesme, D., and Brisset, A., to be published.
19. Adam, J.-C., Laval, G., and Pesme, D., in Proceedings of the International Conference on Plasma Physics, Nagoya, Japan, 1980.
20. O'Neil, T. M., and MalMBERG, J. H., Phys. Fluids 11, 1754 (1968).
21. Bezzerides, B., Gitomer, S. J., and Forslund, D. J., Phys. Rev. Lett. 44, 651 (1980).
22. Drummond, W. E., and Pines, D., Nucl. Fusion, Suppl. Part 3, 1049 (1962). Vedenov, A. A., Velikhov, E. P. and Sagdeev, R. Z., Nucl. Fusion, Suppl. Part 2, 465 (1962).
23. Aamodt, R. E., and Drummond, W. E., Phys. Fluids 7, 1816 (1964). Zaslavsky, G. M., and Chirikov, B. V., Usp. Fiz. Nauk 14, 195 (1972) [Sov. Phys. Usp. 14, 549 (1972)].
24. O'Neil, T. M., Phys. Rev. Lett. 33, 73 (1974).
25. Adam, J. C., Laval, G., and Pesme, D., Phys. Rev. Lett. 43, 1671 (1979).
26. Laval, G., and Pesme, D., Phys. Letters 80A, 266 (1980).
27. Rudakov, L. I., and Tsytovich, V. N., Plasma Phys. 13, 213 (1971). Choi, D., and Horton, W. Jr., Phys. Fluids 17, 2048 (1974).
28. Adam, J. C., Laval, G., and Pesme, D., Suppl. J. de Physique 41, C3-383, (1980).
29. DuBois, D. F., Phys. Rev. A 23, 865 (1981).
30. Kadomtsev, B. B., Plasma Turbulence (Academic Press, New York, 1965).
31. Kraichnan, R. H., J. Math. Phys. 2, 124 (1961).
32. Orszag, S. A., and Kraichnan, R. H., Phys. Fluids 10, 1720 (1967).
33. DuBois, D. F., and Espedal, M., Plasma Phys. 20, 1209 (1978).
34. Krommes, J. A., and Kleva, R. G., Phys. Fluids 22, 2168 (1979).
35. Krommes, J. A., and Similon, P., Phys. Fluids 23, 1553 (1980).
36. Martin, P. C., Siggia, E. D., and Rose, H. A., Phys. Rev. A 8, 423 (1973).

37. The condition $\tau_h \ll \tau_{<f>}$ appears to be only a restriction on the level of electric field fluctuations whereas the condition for smallness of the vertex corrections to the DIA must also involve some restriction on $\langle \delta f^2 \rangle$. The fluctuation level $\langle \delta f^2 \rangle$ is asymptotically in time proportional to $\langle \delta E^2 \rangle$ and both are bounded. Convergence would seem possible for sufficiently small and smooth initial fluctuations $\langle \delta f^2 \rangle$.

SELF-FOCUSING TENDENCIES OF THE NONLINEAR
SCHRÖDINGER AND ZAKHAROV EQUATIONS

HARVEY A. ROSE AND D. F. DUBOIS

Los Alamos National Laboratory
P. O. Box 1663
Los Alamos, NM 87545

The Zakharov equations and an associated nonlinear Schrödinger (NLS) equation are models of long wavelength, interacting Langmuir waves. Although there seems to be a parameter regime in which these models consist of weakly interacting waves, this regime is pre-empted by self-focusing; which, in the case of the NLS equation, can generate a singularity in a finite time. The implications for a statistical theory are discussed.

1. INTRODUCTION

The problem of Langmuir turbulence in plasmas is one of the conceptually simplest and most studied problems in nonlinear plasma physics.^{1,2,3} The physical situation is that of a completely ionized, unmagnetized plasma with electron temperature much higher than the ion temperature where the linear collective modes of the system are longitudinal Langmuir modes and ion-sound modes and transverse electromagnetic modes. A useful model which we will use here and which describes many of the interesting non-linear features of the system is the model of Zakharov⁴ which explicitly separates the fast and slow time scales of the system.

Weak turbulence theory is built on the linear collective modes of the system. It has been studied since the early 1960's.² It is well-known^{1,2} that this theory predicts a transfer of Langmuir wave intensity to low wave numbers by a cascade of three wave processes of the form $\ell \rightleftharpoons \ell' + s$, i.e., the three wave process which converts a Langmuir wave (ℓ) with wave number k into a Langmuir wave (ℓ') with number k' and an ion-sound wave with wave number $k - k'$ and vice versa. The frequency matching condition for these processes is the well-known one

$$\omega_k = \omega_{k'} + |k - k'|c_s \quad (1.1)$$

where $\omega_k = (\omega_p^2 + 3k^2v_e^2)^{1/2}$ is the Langmuir frequency and $c_s = (T_e/m_i)^{1/2}$ is the ion-sound speed where we express temperature in energy units. If Langmuir energy is injected at some finite $k \gg (1/3)(m_e/m_i)^{1/2}k_{De} = k^*$ then Langmuir intensity cascades to lower $k' < k$ until i.) the intensity is low enough for some finite $k_{min} > k^*$ so that the mode k_{min} is stabilized by linear collisional damping or ii.) unstable modes for $k < k^*$ are excited for which further decay by the process (1.1) are not allowed. This piling up of Langmuir intensity at low k has been called the "Langmuir condensate." The question then arises as to how the energy in this condensate is dissipated. Attempts have been made to invoke weak turbulence processes such as wave-wave scattering, i.e., $\ell_1 + \ell_2 = \ell_3 + \ell_4$. However, as we will show here explicitly the weak turbulence scheme breaks down for low wavenumbers $k < k^*$.

It is known^{3,4,1} that coherent or incoherent Langmuir oscillations at long wavelengths are unstable to modulational instability in which regions of excess Langmuir oscillations deplete the local electron and ion density by action of the ponderomotive force, and the density depletion through its effect on the Langmuir wave phase speed, attracts the trajectories of nearby waves. This instability involves the competition of linear dispersive and dissipative terms with the nonlinear "self-focusing" effect of the ponderomotive force; this is a case of effective "Reynold's number" of order unity. The linear stage of the modulational instability does not directly transfer energy to the high k region where Landau damping can be an effective dissipation process.

The conceptual picture of the nonlinear evolution of the modulational instability derives from a combination of computer experiments and special analytic solutions.¹ In the case of one dimension, it appears^{5,6} that the modulational instability can develop into a set of spiky structures which may be loosely related to the known soliton initial value solutions of the undriven, undamped Zakharov or nonlinear Schrödinger equations. When the system is driven, e.g., by a uniform pump, there is conflicting evidence⁵ that these spiky structures may, in fact, undergo a singular collapse. In two or three dimensions it is well-known from the pioneering work of Zakharov⁴ and more recent work by Goldman and Nicholson⁶ that a localized Langmuir wavepacket will tend to collapse if the electrostatic energy in the packet W satisfies

$$\frac{W}{nT_e} > f \left(\frac{\Delta k}{k_D} \right)^2$$

where Δk is the width of the spectrum in wave number and f is a constant of order unity. The ultimate evolution of the collapsing wave packet presumably involves strong wave particle interaction in which the energy of the packet is given up to very energetic electrons. In two or three dimensions it appears that direct collapse can be initiated without the intermediate stage of modulational instability.

It is conjectured that the nonlinear stage of the modulational instability results in small scale structures of the order of several Debye lengths which dissipate energy by wave particle interaction. Galeev, et al.⁷ argue that small scale density fluctuations are created by sound radiation from supersonically collapsing wave packets; these small scale density fluctuations provide an enhanced absorption (analogous to inverse Bremsstrahlung) of longer wavelength Langmuir waves which terminates the collapse.

In our research we have attempted to formulate a systematic statistical theory of Langmuir turbulence using the "strong" turbulence or renormalization methods of Kraichnan^{8,9} which were originally applied to incompressible fluid turbulence. The simplest renormalized theory, the so-called Direct Interaction Approximation (DIA) has been successful in describing moderate Reynold's number fluid turbulence so it seemed reasonable to apply it to the Langmuir turbulence problem with "Reynold's number" of order unity.

In applying these methods to the Langmuir turbulence problem one is immediately faced with problems not encountered in the usual applications to isotropic, homogeneous Navier-Stokes turbulence. Among these problems is the existence of weakly damped, oscillatory modes corresponding in the limit of large k to the unperturbed (Langmuir and ion sound) modes of the system and the existence of nonlinear normal modes due to self focusing, i.e., modulational instability which goes over to transient self trapping in the steady state. These problems complicate both the numerical solution of the DIA equations and further analytic approximations such as the "pole approximation" discussed below.

2. THE ZAKHAROV EQUATIONS

Let $\psi(\underline{x}, t)$ be the (complex valued) low frequency envelope of the electrostatic potential. Then the Zakharov equations are ^{4,1,2}

$$\nabla^2 \left[i \left(\frac{\partial}{\partial t} + \nu_e \right) + \nabla^2 \right] \psi(\underline{x}, t) = \nabla \cdot [n(\underline{x}, t) \nabla \psi(\underline{x}, t)] \quad (2.1)$$

$$\left(\frac{\partial^2}{\partial t^2} + 2\nu_i \frac{\partial}{\partial t} - \nabla^2 \right) n(\underline{x}, t) = \nabla^2 [\nabla \psi(\underline{x}, t) \cdot \nabla \psi^*(\underline{x}, t)] \quad (2.2)$$

where $\psi^*(\underline{x}, t)$ is the complex conjugate of $\psi(\underline{x}, t)$.

We are using a dimensionless set of units here¹ in which distances are measured in units of $\frac{3}{2}(m_i/m_e)^{1/2} \lambda_{De}$ where λ_{De} is the electron Debye length, time is measured in units of $\frac{3}{2}(m_i/m_e) \omega_{pe}^{-1}$ where ω_{pe} is the electron plasma frequency, electric field strengths are measured in units of $(m_e/m_i)^{1/2} (16\pi n_0 T_e/3)^{1/3}$ where n_0 is the average ion or electron density and T_e is the electron temperature and the density fluctuation n is measured in units of $n_0 (4m_e/3m_i)$. Phenomenological Langmuir wave damping ν_e and ion wave damping ν_i (in the above units) have been added to the equations; ν_e can be chosen to be nonlocal in real space to mimic Landau damping if desired. We will not discuss the conditions for validity of these equations here since they have been given elsewhere.¹ Our goal is to study the statistical theory of these model nonlinear equations; our hope is that further physical refinements can be added later in a straightforward way. They can be cast into Hamiltonian form¹ when the dissipative coefficients, ν_e and ν_i , vanish by formally regarding ψ and ψ^* as independent fields and by introducing the hydrodynamic flux potential u :

$$\nabla^2 \left[i \left(\frac{\partial}{\partial t} + \nu_e \right) + \nabla^2 \right] \psi = \nabla \cdot [n \nabla \psi] \quad (2.3a)$$

$$\nabla^2 \left[-i \left(\frac{\partial}{\partial t} + \nu_e \right) + \nabla^2 \right] \psi^* = \nabla \cdot [n \nabla \psi^*] \quad (2.3b)$$

$$\frac{\partial n}{\partial t} = \nabla^2 u \quad (2.3c)$$

$$\left(\frac{\partial}{\partial t} + 2v_i\right)u = n + \nabla\psi \cdot \nabla\psi^* \quad (2.3d)$$

Define the Hamiltonian H as the volume integral of the density h .

$$h = \nabla^2\psi \nabla^2\psi^* + \frac{1}{2}[(\nabla u)^2 + n^2] + n \nabla\psi \cdot \nabla\psi^* \quad (2.4)$$

Then equations (2) are equivalent, when $v_e = v_i = 0$, to the canonical system.

$$\frac{\partial n}{\partial t} = -\delta H / \delta u \qquad \frac{\partial u}{\partial t} = \delta H / \delta n \quad (2.5)$$

$$\nabla^2 \frac{\partial \psi}{\partial t} = i \delta H / \delta \psi^* \qquad \nabla^2 \frac{\partial \psi^*}{\partial t} = -i \delta H / \delta \psi$$

In addition to H , equations (2.5) conserve: the plasma number N ,

$$N = \int \nabla\psi \cdot \nabla\psi^* d\mathbf{x} = \int \mathbf{E} \cdot \mathbf{E}^* d\mathbf{x} \quad (2.6)$$

where \mathbf{E} is the complex electric field envelope, $-\nabla\psi$; the linear momentum

$$P_L = \int \left\{ \frac{1}{2} i (E_m \partial E_L^* / \partial x_m - E_m^* \partial E_L / \partial x_m) - n \partial u / \partial x_L \right\} d\mathbf{x} \quad (2.7)$$

and the angular momentum

$$\mathbf{M} = \int \{ i \mathbf{E} \times \mathbf{E}^* + \mathbf{x} \times \mathbf{p} \} d\mathbf{x} \quad (2.8)$$

where \mathbf{p} is the linear momentum density.

3. THE NONLINEAR SCHRÖDINGER EQUATION

The NLS equation;

$$\frac{\partial \underline{E}}{\partial t} + \frac{1}{2} \nabla (\nabla \cdot \underline{E}) + |\underline{E}|^2 \underline{E} = 0 \quad (3.1)$$

is derivable from the Hamiltonian density

$$h = \frac{1}{2} [|\nabla \cdot \underline{E}|^2 - |\underline{E}|^4] \quad (3.2)$$

It has the plasmon number as an invariant, as well as momentum

$$\underline{P} = \frac{1}{2i} \int [\underline{E}^* \nabla \cdot \underline{E} - \underline{E} \nabla \cdot \underline{E}^*] d\mathbf{x} \equiv \int p(\mathbf{x}) d\mathbf{x} \quad (3.3)$$

In special dimension, d , equal to one, it has an infinite number of invariants. There is a close relation between (3.1) and the adiabatic limit of (2.1), (2.2): at long wavelengths ($\gg 1$) the fast phase speed of the ion sound waves enables the density fluctuations to adjust rapidly to the instantaneous ponderomotive force so that (2.2) can be replaced by

$$n = -|\nabla \psi|^2. \quad (3.3)$$

When this is substituted into (2.1) a NLS equation is obtained which resembles (3.1), but for which there is no self focusing theorem with the generality of (5.1).

4. FAILURE OF A WEAKLY INTERACTING WAVE APPROXIMATION

If the electric field has a small amplitude, can the $|\underline{E}|^4$ term in (3.2) be treated as a correction to H_0 ,

$$H_0 = \frac{1}{2} \int |\nabla \cdot \mathbf{E}|^2 d\mathbf{x}, \quad H = H_0 + H_I \quad ?$$

Even if this is true initially, the evolution of (3.1) can lead to a spectral distribution [of the Fourier transform of \mathbf{E} , $\mathbf{E}(\mathbf{k})$] which violates $|H_I| \ll H_0$. The general tendencies for spectral evolution can be inferred from the conservation laws for H and N . Let us define the spectral function

$$C(\mathbf{k}) = k^{d-1} \int |\mathbf{E}(\mathbf{k})|^2 d\Omega$$

where $d\Omega$ is the solid angle differential. If $|H_I| \ll H_0$, then

$$H \approx H_0 = \int_0^\infty k^2 C(\mathbf{k}) dk,$$

and, in general,

$$N = \int_0^\infty C(\mathbf{k}) dk.$$

If, at $t = 0$, $C(\mathbf{k})$ and $k^2 C(\mathbf{k})$ are peaked at a wave number k_0 , the evolution of the NLS equation cannot lead to a spectral function similarly peaked at a $k \neq k_0$ because the area under the peak must be constant to conserve N , while its k^2 moment must be constant to conserve H_0 . Nor can there be a symmetric broadening of the peak since this would increase H_0 . A depletion of the peak with a stronger transfer of $C(\mathbf{k})$ towards $k < k_0$ is the only kind of spectral solution consistent with conservation of H_0 and N .

In the framework of weakly interacting waves, a kinetic equation for the evolution of $C(\mathbf{k}, t)$, whose collision term only includes resonant wave interactions, is known as weak turbulence theory (WTT). The scaling solutions ($C(\mathbf{k}) \sim 1/k^n$) for the NLS equation have been discussed by Zakharov^{4,10} who showed that these solutions have a flux of N towards the infrared, and a flux of H_0 towards the ultraviolet. For the Zakharov equations, this tendency has been seen in numerical simulations.⁶ These are two confirmations of the tendency deduced from general principles in the preceding paragraph. (A similar general tendency was pointed out by Kraichnan¹¹ for the two dimensional Navier Stokes equation.)

This transfer of N to ever longer wavelengths, will eventually lead to a spectral distribution which is inconsistent with the ansatz $|H_I| \ll H_0$. This follows from the gradients (k^2 factor) in H_0 , and the absence of gradients in H_I . One specific mechanism which will come into play is the modulational instability. This instability is easily seen in the special case of a spacially uniform solution to the NLS equation (i.e. a spectrum peaked at $k = 0$, with a peak width of zero). Aside from a time dependent phase factor, this class of solutions is characterized by $|E|^2 = W = N/\text{unit length}$

$$E(x, t) = E e^{iWt}$$

The linearization of the NLS equation about this solution yields an eigenvalue problem, with eigenfunctions, δE , of the form $\delta E \sim e^{ikx}$. The eigenvalue has an imaginary part, γ , for $k^2 < 2W$ (we specialize here to $d = 1$)

$$\gamma = k(2W - k^2)^{1/2}.$$

If the spectrum has a wave number width Δ , then for Δ small enough so that there is a modulational instability,

$$\gamma = k[(2W - k^2)^{1/2} - A \cdot \Delta]$$

where A is a constant of order unity.

5. COLLAPSING SOLUTIONS OF THE NLS EQUATION

Goldman and Nicholson⁶ have proven the following theorem for solutions to the NLS equation, (3.1). If the solution is spacially localized, so that various surface integrals at infinity can be ignored, then

$$\begin{aligned} \overline{(\delta r)^2} = [2H/N - P^2/N^2]t^2 + Bt + \overline{(\delta r)^2}_{t=0} \\ + (2-d) \int_0^t dt' \int_0^{t'} dt'' \overline{|E|^2} \end{aligned} \quad (5.1)$$

where d is the spacial dimension,

$$\bar{f} = \int d\mathbf{r} f(\mathbf{r}, t) |E|^2(\mathbf{r}, t)/N \quad (5.2)$$

$$\delta \mathbf{r} = \mathbf{r} - \bar{\mathbf{r}}$$

$$B = \left(\frac{-i}{N}\right) \int d\mathbf{r} \delta \mathbf{r}(t=0) \cdot [\mathbf{E}^* \nabla \cdot \mathbf{E} - \mathbf{E} \nabla \cdot \mathbf{E}^*]_{t=0} \quad (5.3)$$

This implies that if

$$H < P^2/2N \quad (5.4)$$

then for $d > 2$

$$\overline{\delta r^2}|_t = t_c = 0 \quad (5.5)$$

where t_c is the collapse time, $0 < t_c < \infty$ unless the solution $E(\mathbf{x}, t)$ develops a singularity at some time $t < t_c$. If the first singularity occurs at t_c , then for $t \rightarrow t_c$, $|E|$ is going to zero everywhere, except at $\mathbf{r} = \bar{\mathbf{r}}(t)$, where $|E| \rightarrow \infty$ because N is a constant. This is aptly described as "collapse."

Unless the initial conditions have a special symmetry, we believe that t_c is pre-empted by a less violent kind of singularity. Consider an initial condition, E_0 , which is the superposition of two fields, E'_0 and E''_0 . each of the latter two are spherically symmetric, spacially localized and identical to each other except that one is centered at $\mathbf{r} = (-L, 0, 0)$, and the other at $(L, 0, 0)$. As $L \rightarrow \infty$, it is plausible that these two spheres evolve as if its twin were absent, especially if each sphere by itself satisfies $H < P^2/2N$. Therefore the time for the formation of a singularity should be the same as for a single sphere initial condition. However, the initial value of $\overline{\delta r^2}$ is the only parameter in (5.1) which approaches infinity as L does (it goes like L^2). Therefore (5.5) implies

$$\lim_{L \rightarrow \infty} t_c \sim L \quad .$$

The basic reason why t_c is hardly ever attained is that the NLS equation is spacially local, while the collapse condition (5.4) is a global property. Since the following transformation takes solutions of the NLS equation into another solution:

$$E(x,t) \rightarrow S \cdot E(Sx, S^2t) \quad (5.6)$$

where S is a scale factor, there are only intrinsic measures of distance, one of which is determined by the field strength

$$l_E = 1/|E(x,t)|$$

and another is the gradient scale length

$$l_V = |E|^{-1} |\nabla E|^{-1}.$$

Perhaps the collapse condition (5.4) can be improved by replacing the global integrals of $h(x)$, $p(x)$, and $|E|^2(x)$ (to give H , P , and N) by local integrals of linear extent l_E or l_V . The condition obtained by using l_V is consistent with our intuition concerning the two, well separated, sphere initial condition.

Of course, the physics of plasmas would invalidate the NLS equation if field strengths and gradients become too large. However, it is conceivable that with initial plasma conditions which were characterized by very weak, long wavelength fields (field energy \ll thermal energy, gradient scale lengths \gg Debye radius) the plasma could evolve to a state which is strongly collapsed relative to the initial conditions, but still be in a physical regime where the NLS equation is still valid.

As collapse continues in time the density will not keep up with the rapid variations of E , and the adiabatic relation (3.3) must be replaced by the Zakharov equations. These too will become invalid as collapse intensifies, and to properly account for the interaction between fields and particles the Vlasov equation is needed.

If collapse did occur for solutions of the NLS equation, in the strict sense that $(\delta r)^2 \rightarrow 0$, then the singularity would be at a point, i.e., a set with dimension zero. Even though this may not be realized, it still must be true that whatever the nature of the singularity, infinite field strengths are present because otherwise the dispersive term in (3.1) would dominate the nonlinear term, and the resulting linear Schrödinger equation is well behaved. Since N is conserved, the regions where $|E|$ is becoming stronger and stronger must fill a smaller and smaller fraction of the spacial volume, and this is what we mean by intermittency.

It is unlikely that the statistical theories of the kind discussed in the next section can accurately describe a strongly intermittent solution of the NLS equation. However, there are physically relevant parameters regimes in which collapse is cut off by dissipative effects so that the intermittency of the solutions to the highly idealized NLS equation is not actually realized. Still, there is a tendency to self-focus, so that a theory that is more comprehensive than weak turbulence theory is called for.

6. STATISTICAL THEORY OF THE ZAKHAROV AND NLS EQUATIONS

Since the NLS equation is a long wavelength limit of the Zakharov equations, it is possible to develop a statistical theory of the former by taking the long wavelength limit of the latter. Conversely since, by solving (2.2) for n in terms of $|\nabla\psi|^2$, and substituting in (2.1), the Zakharov equation for ψ has the form of a generalized NLS equation, a statistical theory of a NLS equation could be used for the Zakharov equations.

It has been found that the DIA¹² as applied to the Zakharov equations is essentially equivalent¹³ to the Screened Potential Approximation for the NLS equation (i.e., the classical field theory analog of the quantum mechanical plasma approximation that is needed to understand electrodynamic screening of charged particles). The propagation of infinitesimal disturbances in the density field n , R^n , of (2.2) (disturbances of $|E|^2$ in 3.1) is modified by polarization-like effects, which for the Zakharov equation DIA, amounts to replacing k^2 , the fourier transform of $(-\partial^2/\partial x^2)$ in the inverse of the sound wave propagator,

$$\frac{\partial^2}{\partial t^2} + 2v_i \frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}$$

by

$$k^2 + k^2 \sigma(k, \omega)$$

where $\sigma(k, \omega)$ is the transform of $\sigma(x, t)$,

$$\sigma(x, t) = i[R^\psi(xt)C^*(xt) - \text{c.c.}]$$

and where R^ψ is the mean propagator of infinitesimal disturbances in ψ , and

$$C(xt) = \langle \partial \psi(y\tau) / \partial y \partial \psi^*(y'\tau') / \partial y' \rangle$$

with $y - y' = x$, $\tau - \tau' = t$. Homogeneous turbulence has been assumed and the results presented for $d = 1$, with similar results for $d > 1$.

While the traditional polarization of charged particles involves a rearrangement of charge density over an infinite distance (i.e., Debye screening) when a local charge inhomogeneity lasts a long time, in the Zakharov and NLS equations, there can be transient but long lived (or even unstable) trapping of Langmuir waves by density inhomogeneities if the wave number bandwidth of the spectrum of fluctuations of E is small compared to $(W)^{1/2}$. These are the self-focusing and modulational instability mechanisms mentioned earlier.

For the special case of a narrow bandwidth spectrum of Langmuir wave fluctuations centered at $k = 0$, we have compared the linear response functions as predicted by the DIA with a Monte Carlo simulation of a few Fourier modes extracted from the Zakharov equations and found qualitative agreement.¹²

As an aid in understanding the structure of the DIA, and also for the purpose of finding a method of solution that is more efficient than numerical integration of the DIA, we¹² have developed a consolidated continued fraction technique that is applied to the space and time Fourier transformed correlation and linear response functions $C(k\omega)$ and $R(k\omega)$. Let us recall that the form of the DIA is

$$G^{-1} = G_0^{-1} - \mathcal{J}(G) \quad (5.1)$$

where G is a generalized propagator that includes both C and R ; and \mathcal{J} is a quadratic functional. Equation (5.1) is replaced by

$$G_{\ell+1}^{-1} = G_0^{-1} - \mathcal{J}(G_\ell) \quad \ell = 0, 1, \dots \quad (5.2)$$

Under some mild restrictions it is easy to show that each iterate, $G_\ell(k\omega)$ is a rational function of ω and hence is characterized by a number, $n(\ell)$, of poles and residues. Since $n(\ell)$ is a rapidly growing function of ℓ , a numerical investigation of the possible convergence of (5.2) must involve some kind of consolidation of the poles at each iteration. For the particular cases that we have examined, this technique (using rough, ad hoc methods of consolidation) appears to succeed in the sense that convergence is frequently obtained, with the fixed point solution to (5.2) in good agreement with the solution to the DIA, determined by numerical integration.

It is natural to interpret the collection of poles of $G(k\omega)$ for fixed k , as nonlinear normal modes of the turbulent system because they depend on the strength and spectral distribution of fluctuations. In addition to modes that correspond to turbulently modified continuations of the linear normal modes (these are the only ones allowed by the various weak turbulence theories) there are new modes that correspond to the dynamics of self-focusing and modulational instability.

Although this technique is still numerically complex, it should serve as an intermediary in the construction of a simpler phenomenological theory at the level of the Eddy damped quasi normal approximations¹⁴ that have proved to be of great utility in the study of fully developed fluid turbulence.

REFERENCES

1. Two useful review articles on the subject of Langmuir turbulence are: S. G. Thornhill and D. ter Haar, "Langmuir Turbulence and Modulational Instability", *Physics Reports* **43C** (1978) 43 and L. I. Rudakov and V. N. Tsytovich, "Strong Langmuir Turbulence", *Physics Reports* **40** (1978) 1.

2. B. B. Kadomstev, "Plasma Turbulence", Academic Press, New York, 1965.
V. N. Tsytovich, "Theory of Turbulent Plasma", Consultants Bureau, New York, 1977.
3. A. A. Vedenov and L. I. Rudakov, Sov. Phys. Doklady 9 (1965) 1073.
4. V. E. Zakharov, Sov. Phys. JETP 35 (1972) 908.
5. J. J. Thomson, F. J. Faehl, and W. L. Kruer, Phys. Rev. Lett. 31 (1973) 918. G. J. Morales, Y. C. Lee and R. B. White, Phys. Rev. Lett. 32 (1974) 457.
6. M. V. Goldman and D. W. Nicholson, Phys. Rev. Lett. 41 (1979) 406.
D. R. Nicholson, M. V. Goldman, P. Hoyne, and J. C. Weatherall, Astrophysical J. 223 (1978) 605. D. R. Nicholson and M. V. Goldman, Phys. Fluids 21 (1978) 1766.
7. A. A. Galeev, R. Z. Sagdeev, V. D. Shapiro, and V. I. Schevchenko, JETP Letters 24 (1976) 21.
8. R. H. Kraichnan, J. Fluid Mech. 3 (1959) 32.
9. R. H. Kraichnan, Phys. Fluids 7 (1964) 1163.
10. V. E. Zakharov and E. A. Kuznetsov, JETP 48 (1978) 458.
11. R. H. Kraichnan, Phys. Fluids 10 (1967) 1417.
12. D. F. DuBois and H. A. Rose, "Statistical Theories of Langmuir Turbulence: I. The Direct Interaction Approximation Responses, Los Alamos report LA-UR-80-3603 (accepted for publication in Phys. Rev. A).
13. D. F. Dubois, D. R. Nicholson, and H. A. Rose, "Statistical Theories of Langmuir Turbulence II (in preparation).
14. For a review of these kinds of turbulence approximations see: H. A. Rose and P. L. Sulem, Journal de Physique 39 (1978) 441.

CHAOTIC OSCILLATIONS IN A SIMPLIFIED MODEL FOR LANGMUIR WAVES

P. K. C. Wang

School of Engineering and Applied Science
University of California
Los Angeles, California
U.S.A.

The results of a further study of the chaotic solutions of the single-mode equations corresponding to Zakharov's model for Langmuir waves in a plasma are described briefly.

Recently, it was found [1] that the following single-mode equations corresponding to Zakharov's model for Langmuir waves in a plasma have nonperiodic chaotic solutions

$$i \frac{dE_m}{dt} - (\mu_m - i\gamma_m)E_m = n_m(E_0 + \alpha_m E_m), \quad (1)$$

$$\frac{d^2 n_m}{dt^2} + 2\Gamma_m \frac{dn_m}{dt} + \mu_m^2 n_m = -\mu_{m0}^2 E_0 \cdot (E_m + E_m^*) + \beta_m |E_m|^2, \quad (2)$$

where

$$\alpha_m = \int_{\Omega} \phi_m^3(\underline{x}) d\Omega, \quad \beta_m = \int_{\Omega} \nabla^2 [\phi_m^2(\underline{x})] \phi_m(\underline{x}) d\Omega. \quad (3)$$

Here, $\underline{a} \cdot \underline{b}$ denotes the usual scalar product of two vectors \underline{a} and \underline{b} in the real Euclidean space \mathbb{R}^N ; ϕ_m is the orthonormalized eigenfunction of the Laplacian operator corresponding to the eigenvalue $-\mu_m < 0$; γ_m and Γ_m are the phenomenological damping coefficients; E_m and n_m are respectively the coefficients of expansions for the electric field \underline{E} and ion density n in terms of ϕ_m defined on a bounded spatial domain $\Omega \subset \mathbb{R}^N$. $E_0 \in \mathbb{R}^N$ represents the normalized amplitude of an external spatially homogeneous electric field (pump) oscillating at the electron plasma frequency. In [1], chaotic solutions were discovered through numerical experimentation and bifurcation analysis with $\|E_0\|^2$ as a variable parameter. The structure and the onset of these solutions were not explored in detail. Here, we summarize the results of a further study of the chaotic solutions of (1)-(3).

1. Bifurcation of Equilibrium States

For fixed $\delta_m > 0$ and variable parameter $\|E_0\|^2$, the trivial solution $\underline{z} = (n_m, \dot{n}_m, E_m) = (0, 0, 0 + i0)$ is the only equilibrium state when $0 \leq \|E_0\|^2 < E_{0c}^2$, where

$$E_{0c}^2 \triangleq 2\alpha_m \delta_m^{-2} [\beta_m + (\beta_m^2 + \delta_m^2 \gamma_m^2)^{\frac{1}{2}}], \quad \delta_m \triangleq (2\alpha_m + \beta_m \mu_m^{-2}). \quad (4)$$

At $\|E_0\|^2 = E_{0c}^2$, a nontrivial equilibrium state z_e emerges and it bifurcates into two distinct equilibrium states z_e^+ and z_e^- when $\|E_0\|^2 > E_{0c}^2$. For the case where $\|E_0\|^2$ is fixed and γ_m is a variable parameter, the trivial solution is the only equilibrium state when $\gamma_m = 0$. For $0 < \gamma_m < \gamma_m^c$ and $\gamma_m \neq \gamma_m^c$, where

$$\gamma_m^c \triangleq \|E_0\| \{ \|E_0\|^2 (2\alpha_m \mu_m^2 + \beta_m)^2 - 4\alpha_m \beta_m \mu_m^4 \}^{\frac{1}{2}} / (2\alpha_m \mu_m^2),$$

$$\gamma_m^c \triangleq \mu_m [2\|E_0\|^2 - \mu_m^2]^{\frac{1}{2}}, \quad (5)$$

we have, in addition to the trivial equilibrium state, two distinct nontrivial equilibrium states and they merge into one when $\gamma_m = \gamma_m^c$. When γ_m increases beyond γ_m^c , the nontrivial equilibrium states abruptly disappear. At $\gamma_m = \gamma_m^c$, only a trivial and a nontrivial equilibrium states exist.

2. Chaotic Oscillations

In the case where γ_m and Γ_m correspond to Landau damping, γ_m varies rapidly for small $\mu_m \lambda_D > 0$, where λ_D is the Debye length. It is of interest to determine the threshold value of γ_m for the quenching or onset of chaotic oscillations. A detailed numerical study was made for the one dimensional case with $\Omega =]0, L[$, $\phi_1(x) = \sqrt{2/L} \sin(\pi x/L)$, $\mu_1 = \pi/L$, $L = \pi/\sqrt{10}$, $\Gamma_1 = 2$, $\alpha_1 = \sqrt{(2/L)/(3\pi)}$ and $\beta_1 = -8\pi\sqrt{(2/L^5)}/3$. For the typical case where $E_0^2 = 2.669$, we have $\gamma_1^c = 5.3358$. It can be verified that Hopf bifurcation (with respect to the parameter $\hat{\gamma} = 1/\gamma_1$) takes place about the equilibrium state z_e^+ when $\gamma_1 = \gamma_H \approx 4.35$. As γ_1 is decreased from γ_H , it was found that the solutions about z_e^+ change from periodic to almost periodic and become chaotic when $\gamma_1 = \gamma_T \approx 1.55$. The power spectra of the electric field evolve from essentially discrete spectra to broadband spectra as γ_1 crosses γ_T . The phenomena appear to be more complex than simple period doubling as observed in many systems. We also obtained numerically the points generated by Poincaré maps corresponding to various hyperplanes in z -space which are transverse to the chaotic trajectories. Typical results are shown in Figures 1 and 2. Evidently, the points appear to lie in the neighborhood of some curve in \mathbb{R}^3 , implying that the Poincaré map is essentially one-dimensional in nature. However, the graphs of the Poincaré maps in terms of some curve parameter are not readily obtainable. A more detailed description of the above-mentioned results is given in reference [2].

REFERENCES

- [1] Wang, P.K.C., "Nonlinear Oscillations of Langmuir Waves", J. Math. Physics, 21 (1980) 398-407.
- [2] Wang, P.K.C., "Further Study of Chaotic Oscillations of Simplified Zakharov's Model for Langmuir Turbulence", Univ. of Calif. Engr. Rprt. No. UCLA-ENG-8013 (March, 1980).

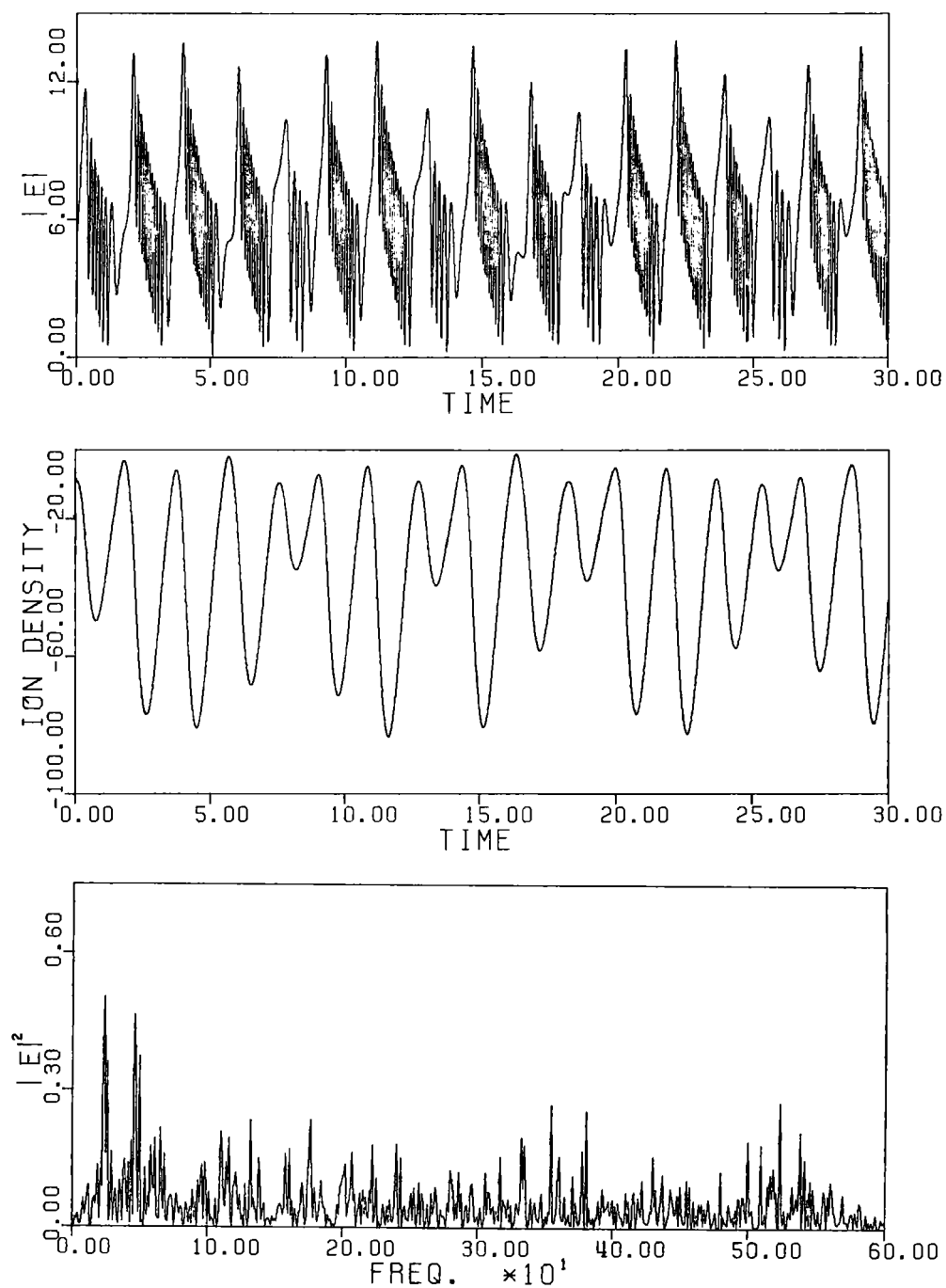


Fig.1: Typical chaotic solution and its corresponding electric field power spectrum ($\gamma_1 = 1.0$, $\Gamma_1 = 2.0$, $E_o^2 = 10$)

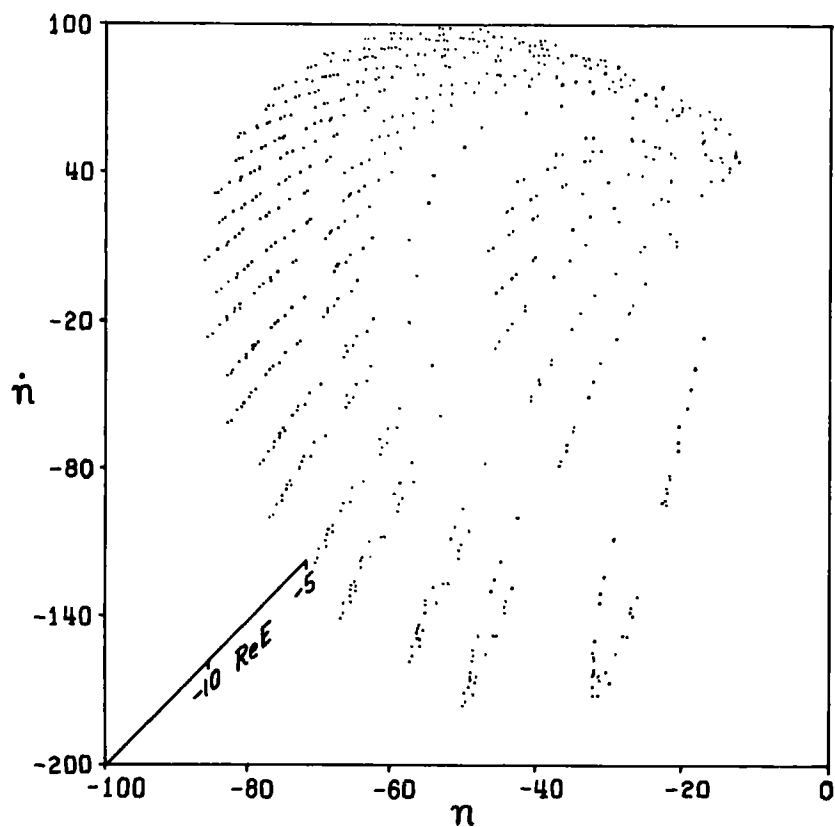


Fig.2: Isometric representation of points generated by Poincaré map (for the solution shown in Fig.1) onto the hyperplane: $\text{Im } E = 0$ in z -space.

INDEX OF CONTRIBUTORS

- Andrei, N., 169
 Aris, R., 247
 Bardos, C., 289
 Bishop, A.R., 195
 Brauner, C.M., 335
 Campbell, D.K., 195
 Chang, S.-J., 395
 Crandall, M.G., 117
 Dubois, D.F., 425, 465
 Emery, V.J., 159
 Etemad, S., 209
 Feigenbaum, M.J., 379
 Fife, P.C., 267
 Flaschka, H., 65
 Foias, C., 317
 Fremond, M., 109
 Friaa, A., 109
 Gollub, J.P., 403
 Greene, J.M., 423
 Heeger, A.J., 209
 Holland, R.R., 229
 Howes, F.A., 329
 Hyman, J.M., 91
 Jeandel, D., 335
 Kumar, P., 229
 Leach, P.G.L., 133
 Lewis, H.R., 133
 Lions, J.-L., 3
 Lions, P.L., 17
 Mathieu, J., 335
 McCormick, W.D., 409
 Newell, A.C., 65
 Newman, W.I., 147
 Orszag, S.A., 367
 Palmore, J.I., 127
 Patera, A.T., 367
 Pesme, D., 435
 Raupp, M., 109
 Rice, M.J., 189
 Rose, H.A., 465
 Roux, J.C., 409
 Ruck, H.M., 237
 Sattinger, D.H., 51
 Schrieffer, J.R., 179
 Singer, I.M., 35
 Swinney, H.L., 409
 Temam, R., 317
 Turner, J.S., 409
 Wang, P.K.C., 479
 Wortis, M., 395
 Wright, J.A., 395

This Page Intentionally Left Blank